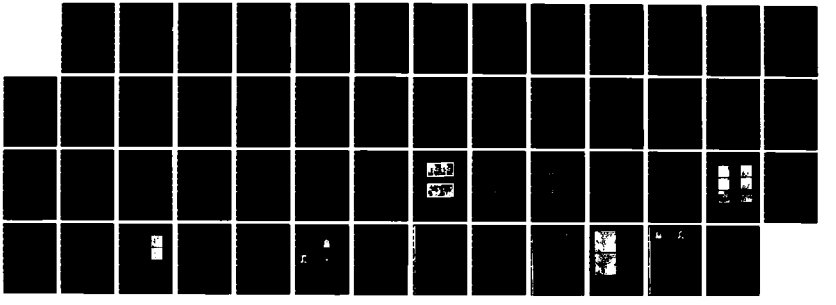
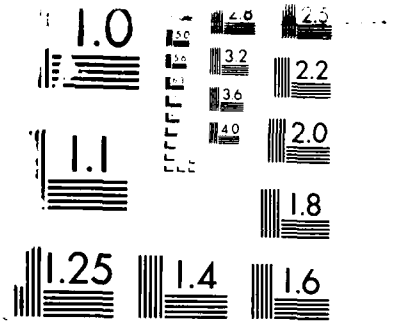
AD-A175 059 STRUCTURE FROM MOTION(U) MINNESOTA UNIV MINNEAPOLIS 1/1
W B THOMPSON DEC 85 AFOSR-TR-86-2132 F49620-83-C-0140

UNCLASSIFIED F/G 20/6 NL




1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0

SECURITY CLASSIFICATI

AD-A175 059

ATION PAGE

1a. REPORT SECURITY C: Unclassified	RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFIC:	3. DISTRIBUTION/AVAILABILITY OF REPORT: Unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE:	
4. PERFORMING ORGANIZATION REPORT NUMBER(S):	5. MONITORING ORGANIZATION REPORT NUMBER(S): AFOSR-TR- 86-2132
6a. NAME OF PERFORMING ORGANIZATION: Universi of Minnesota	6b. OFFICE SYMBOL (If applicable):
6c. ADDRESS (City, State and ZIP Code): 136 Lind Hall, 207 Church St. SE, Minneapolis, MN 55455	7b. ADDRESS (City, State and ZIP Code): Bolling AFB, Washington DC 20332-6448
7a. NAME OF MONITORING ORGANIZATION: AFOSR	
8a. NAME OF FUNDING/SPONSORING ORGANIZATION: AFOSR	8b. OFFICE SYMBOL (If applicable): nm
8c. ADDRESS (City, State and ZIP Code): BH 410 BAFB D.C. 20332-6448	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: F 49 620-83-C-0140
11. TITLE (Include Security Classification): Final Report - Structure From Motion	10. SOURCE OF FUNDING NOS. — PROGRAM ELEMENT NO.: 61102F; PROJECT NO.: 2304; TASK NO.: A5; WORK UNIT NO.:
12. PERSONAL AUTHOR(S): Thompson, William B.	
13a. TYPE OF REPORT: Final	13b. TIME COVERED: FROM 1 July 83 TO 30 Sep 85
14. DATE OF REPORT (Yr., Mo., Day): 1985 - December	15. PAGE COUNT:
16. SUPPLEMENTARY NOTATION:	

17. COSATI CODES		
FIELD	GROUP	SUB. GR.

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number): Image Understanding, Time-Varying Image Analysis, Visual Motion, Optical Flow, Segmentation

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Work on improving gradient-based methods for optical flow estimation has been completed. An understanding of how errors arise makes it possible to define the inherent limitations of the gradient-based technique, obtain estimates of the accuracy of computed values, enhance the performance of the technique, and demonstrate the informative value of some types of errors.

Significant results have been achieved on the problems associated with *motion-based segmentation*. An approach based on understanding the three-dimensional scene structure leading to an edge in optical flow has been developed. As a result, it is possible to simultaneously detect edges and determine important three-dimensional properties of the associated scene surfaces. The methods which have been developed make it possible to distinguish between occluding and occluded surfaces at a boundary. This technique may make it possible to link

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT: UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION: Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL: Dr. Robert N. Buchal	22b. TELEPHONE NUMBER (Include Area Code): (202) 767-4936 — 22c. OFFICE SYMBOL: nm

DD FORM 1473, 83 APR — EDITION OF 1 JAN 73 IS OBSOLETE.

SECURITY CLASSIFICATION OF THIS PAGE

19. (Abstract Continued)

image regions corresponding to a partially occluded object and to produce descriptions of object boundaries that are less affected by occlusion. In addition, being able to distinguish between occluding and occluded boundaries is a crucial step towards determining the three-dimension position of surfaces.

FINAL REPORT - STRUCTURE FROM MOTION
AFOSR Contract F49620-83-0140

a. Objectives.

Our principal objective continues to be the development of a robust computational approach for estimating the spatial organization of a scene using time varying properties of image sequences. Under this contract, we have been investigating improved methods for estimating and interpreting optical flow from image sequences. Emphasis is placed both on what spatial properties should be computed and on appropriate computational architectures for accomplishing this task.

Three related questions have been investigated in this project.

Estimating optical flow.

> What sorts of errors are intrinsic to spatial-temporal gradient techniques for estimating optical flow? The principal objective of this aspect of the work is to develop *a priori* estimates of expected error based on the nature of the actual imagery, and *a posteriori* error estimates as an integral aspect of flow estimation. In addition, the research effort has focused on how flow estimation can be improved based on an understanding of the nature and magnitude of the errors that are likely to arise.

Interpreting optical flow at object boundaries.

> How can the analysis of optical flow be used to detect object boundaries? How can the three-dimensional structure of object boundaries be determined based on optical flow? The principal objective here is to work towards the development of *motion-based segmentation* techniques for image understanding. Motion-based segmentation has the potential not only for locating object boundaries, but also for reducing problems due to occlusion and for providing three-dimensional information useful for object identification and analysis.

Robust methods for determining object motion.

> How can the motion of object relative to the camera be determined in a robust manner? The objective is to categorize the possible motions into a limited number of meaningful classes and to develop methods for recognizing instances of each class.

b. Status of research effort.

Estimating optical flow.

> We have shown that a major difficulty with gradient-based methods is their sensitivity to a number of conditions commonly encountered in real imagery. Highly textured

surfaces, motion boundaries, and depth discontinuities can all be troublesome for gradient-based methods. Fortunately, these problematic areas can be identified in the image. As a part of this contract, we examined the conditions that lead to errors, methods to reduce errors, and the estimation of measurement errors for one class of gradient-based techniques. By understanding how errors arise we are able to define the inherent limitations of the gradient-based technique, obtain estimates of the accuracy of computed values, enhance the performance of the technique, and demonstrate the informative value of some types of errors.

This part of the project has now been completed.

Interpreting optical flow at object boundaries.

Significant results have been achieved on the problems associated with *motion-based segmentation.* Discontinuities in optical flow are necessarily due to surface boundaries or discontinuities in depth in the scene. Thus, detected edges in flow necessarily correspond to important properties of scene geometry, where as edges in properties such as luminance can be due to a wide variety of scene properties. Our approach is based on understanding the three-dimensional scene structure leading to an edge in optical flow. As a result, we can simultaneously detect edges and determine important three-dimensional properties of the associated scene surfaces.

Motion-based segmentation can not only find boundaries that are difficult to locate in a single view, but it can also provide much more information about the structure of the scene. Our approach makes it possible to distinguish between occluding and occluded surfaces at a boundary. Occlusion boundaries arise due to geometric properties of the occluding surface, not the occluded surface. Thus, while the shape of the edge provides significant information on the structure of the occluding surface, it says little or nothing about the structure of the surface being occluded. This technique may make it possible to link image regions corresponding to a partially occluded object and to produce descriptions of object boundaries that are less affected by occlusion. In addition, being able to distinguish between occluding and occluded boundaries is a crucial step towards determining the three-dimension position of surfaces.

Work is continuing on exploiting these results in a variety of image understanding tasks.

Robust methods for determining object motion.

Object motion can be classified based on optical flow into categories that are significant for further interpretation. In our investigations, object motion was divided into four classes: two types of translation and two types or rotation. Complex motions can be described as combinations of these types. The descriptions are qualitative, characterizing the motion in terms of broad classes but not providing precise, qualitative information about trajectories. We have shown that under some circumstances, the categories are detectable using simple differential operations on the optical flow field. Appropriate combinations of detectors can be used to signal motions likely to lead to a collision between the sensor and an object in the field of view. By structuring the technique as a classification operation involving only a limited number of classes, the noise sensitivity

of differential operators can be reduced. For the situations in which the technique is applicable, it is tolerant of noisy, sparse flow fields and requires little information about camera models, motion constraints, or possible objects.

As a result of our research efforts, we discovered that the assumptions required to utilize this approach are not sufficiently realistic. We are currently pursuing alternate approaches to the determination of object motion.

c. Publications.

A. Yonas, M.E. Arterberry, E.A. Opland, and W.B. Thompson, "The effect of surface layout on motion generated depth," Association for Research in Vision and Opthamology (abstract), Sarasota, FL, April 1984.

W.B. Thompson, K.M. Mutch, and V.A. Berzins, "Analyzing Object Motion Based on Optical Flow," *Proc. Seventh International Conference on Pattern Recognition,* July, 1984.

J.K. Kearney and W.B. Thompson, "Gradient-based estimation of optical flow with global optimization," *Proc. 1st Conf. on Artificial Intelligence Applications,* December 1984.

K.M Mutch and W.B. Thompson, "Analysis of accretion and deletion at boundaries in dynamic scenes," *IEEE Trans. Pattern Analysis and Machine Intelligence,* March 1985.

W.B. Thompson, K.M. Mutch, and V.A. Berzins, "Dynamic occlusion analysis in optical flow fields," *IEEE Trans. Pattern Analysis and Machine Intelligence,* July 1985.

J.K. Kearney and W.B. Thompson, "An error analysis of gradient-based methods for optical flow estimation," under review, *IEEE Trans. Pattern Analysis and Machine Intelligence.*

d. Scientific Collaborators.

Research Assistants:

Martha Arterberry
Scott Bolte
Ian Horswill
Joseph Kearney

Ph.D. August 1983. Thesis title: "The Estimation of Optical Flow".

Kathleen Mutch

Ph.D. August 1983. Thesis title: "Qualitative spatial properties of scenes using motion information".

Collaborating Faculty:

Albert Yonas

An Error Analysis of Gradient-Based Methods for Optical Flow Estimation

Joseph K. Kearney

University of Iowa

William B. Thompson
Daniel L. Boley

University of Minnesota

ABSTRACT

Multiple views of a scene can provide important information about the structure and dynamic behavior of three-dimensional objects. To recover this information, it is necessary to estimate *optical flow* – the velocity, on the image, of visible points on object surfaces. One approach for estimating optical flow is based on the relationship between the gradients of image brightness. While gradient-based methods have been widely studied, little attention has been paid to accuracy and reliability of the approach.

We examine the sources of errors in estimates derived from gradient-based techniques. By understanding how errors arise, we are able to define the inherent limitations of the technique, obtain estimates of the accuracy of computed values, enhance the performance of the technique, and demonstrate the informative value of some types of errors.

1. Introduction.

The velocity field that represents the motion of object points across an image is called the optical flow field. Optical flow results from relative motion between a camera and objects in the scene. Methods which estimate optical flow lie within two general classes. *Gradient-based* approaches utilize a relationship between the motion of surfaces and the derivatives of image brightness [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]. *Matching* techniques locate and track small, identifiable regions of the image over time.

For many problems gradient-based methods offer significant advantages over matching techniques. Matching techniques are highly sensitive to ambiguity among the structures to be matched. Optical flow can be accurately estimated for only highly distinguishable regions. This means that flow can only be determined at a sparse sampling of points across the image. Furthermore, it is computationally impractical to estimate matches for a large number of points. The gradient-based approach allows optical flow to be simply computed at a more dense sampling of points than can be obtained with matching methods.

This work was supported by the Air Force Office of Scientific Research contract F49620-83-0140

Index Terms

Vision, Optical Flow, Motion Estimation, Error Analysis, Gradient-Based Methods

Gradient-based techniques avoid the difficult task of finding distinguishable regions or points of interest. The gradient approach leads to algorithms which are characterized by simple computations localized to small regions of the image. These techniques can be applied over the entire image. As we shall see in the analysis that follows, the gradient technique is also sensitive to ambiguous areas -- it is impossible to locally determine the motion of a homogeneous region. However, gradient-based estimates are typically available over a greater area than those obtained reliably by matching. In addition, the loss of precision for gradient-based estimates in ambiguous areas can be quantified. Accuracy measurements can be used to weight the contribution of motion estimates in further analysis or to filter poor estimates from the flow field. These accuracy measurements can be obtained as a by-product of the flow estimation process and require little additional computation.

While gradient-based methods have been widely studied, little attention has been paid to accuracy and reliability of the approach. A major difficulty with gradient-based methods is their sensitivity to conditions commonly encountered in real imagery. Highly textured surfaces, motion boundaries, and depth discontinuities can all be troublesome for gradient-based methods. Fortunately, these problematic areas can be identified in the image. In this paper we examine the conditions that lead to errors, methods to reduce errors, and the estimation of measurement errors for one class of gradient-based techniques. By understanding how errors arise we are able to define the inherent limitations of the gradient-based technique, obtain estimates of the accuracy of computed values, enhance the performance of the technique, and demonstrate the informative value of some types of errors.

2. The Gradient Constraint Equation.

The gradient constraint equation relates optical flow -- velocity on the image (u,v) -- and the image brightness function $I(x,y,t)$. The common assumption of gradient-based techniques is that the observed brightness -- intensity on the image plane -- of any object point is constant over time. Consequently, any change in intensity at a point on the image must be due to motion. Relative motion between the object and camera will cause the position of a point located at point (x,y) at time t to change position on the image over a time interval δt. By the constant brightness assumption, the intensity of the object point will be the same in images sampled at times t and $t+\delta t$. The constant brightness assumption can be formally stated as

$$I(x,y,t) = I(x+\delta x, y+\delta y, t+\delta t). \tag{1}$$

Expanding the image brightness function in a Taylor's series around the point (x,y,t) we obtain

$$I(x+\delta x, y+\delta y, t+\delta t) = I(x,y,t) + \frac{\partial I}{\partial x}\delta x + \frac{\partial I}{\partial y}\delta y + \frac{\partial I}{\partial t}\delta t + h.o.t. \tag{2}$$

A series of simple operations leads to the gradient constraint equation:

$$0 = I_x u + I_y v + I_t \tag{3}$$

where

$$I_x = \frac{\partial I}{\partial x}, \quad I_y = \frac{\partial I}{\partial y}, \quad I_t = \frac{\partial I}{\partial t}$$

A detailed derivation is given by Horn and Schunck [3].

3. Gradient Based Algorithms.

The gradient constraint equation does not by itself provide a means for calculating optical flow. The equation only constrains the values of u and v to lie on a line when plotted in flow coordinates.

The gradient constraint is usually coupled with an assumption that nearby points move in a like manner to arrive at algorithms which solve for optical flow. Groups of neighboring constraint equations are used to collectively constrain the optical flow at a pixel. Constraint lines are combined in one of three ways. Methods of *local optimization* [5, 6, 7, 8, 10] solve a set of constraint lines from a small neighborhood as a system of equations. *Global optimization* [11, 3, 9] techniques minimize an error function based upon the gradient constraint and an assumption of local smoothness of optical flow variations over the entire image. The *clustering* approach [1, 2] operates globally, looking for groups of constraint lines with coinciding points of intersection in flow space.

We will examine the local optimization technique in detail. Although we will not directly address clustering and global optimization, many of the conclusions reached here also apply to these approaches. Another paper examines some implications of this analysis for global optimization methods [12].

4. Local Optimization.

The method of local optimization estimates optical flow by solving a group of gradient constraint lines obtained from a small region of the image as a system of linear equations. Two constraint lines are sufficient to arrive at a unique solution for (u, v). More than two equations may be included in the system to reduce the effects of errors in the constraint lines. The solution to the over-determined system may be found by any of a number of error minimization techniques.

We will examine errors in the solution of two equation systems. In practice one should solve an over-determined system by some method of best fit, such as least squares. The analysis presented here is extended to over-determined systems in [13].

The pair of equations which we will solve to estimate optical flow at point $\mathbf{p}_i = (x_i, y_i, t_i)$ is

$$\begin{aligned} (i) \quad & I_x^{(i)} u + I_y^{(i)} v = -I_t^{(i)} \\ (j) \quad & I_x^{(j)} u + I_y^{(j)} v = -I_t^{(j)} \end{aligned} \tag{4}$$

where the gradients I_x, I_y, and I_t in equations i and j are evaluated at $\mathbf{p}_i$ and a nearby point $\mathbf{p}_j$.

The gradients in the system (4) are estimated from discrete images and will be inaccurate due to noise in the imaging process and sampling measurement error. Also, the values of (u, v) at $\mathbf{p}_i$ and $\mathbf{p}_j$ are assumed to be the same. The formulation will be incorrect to the extent that optical flow differs between the two points. We will examine how gradient estimation error and error resulting from non-constant optical flow leads to errors in the estimated flow vector.

4.1. Gradient Measurement Error.

The estimates of the intensity gradients I_x, I_y, and I_t will be corrupted by errors in the brightness estimates and inaccuracies introduced by sampling the brightness function discretely in time and space. The error in the brightness function is random and results from a variety of sources such as channel noise and quantization of brightness levels. We assume that the brightness error is approximately additive and independent among neighboring pixels. The gradient, estimated from changes in the brightness estimates, will contain a component of random error which is distributed like the error in the brightness function. The random component of the gradient error will be additive and independent of the magnitude of the gradient to the extent that the brightness noise is additive.

The brightness function is sampled discretely in time and space and this will introduce a systematic measurement error into the estimates $\hat{I}_x, \hat{I}_y$, and $\hat{I}_t$ of the gradients. The gradient sampling error depends on the second and higher derivatives of the brightness function. To examine the sampling error in $\hat{I}_x$ we expand the brightness function evaluated at $(x+\Delta x, y, t)$ around the point (x, y, t) producing

$$I(x+\Delta x, y, t) = I(x, y, t) + I_x \Delta x + I_{xx} \Delta x^2 + h.o.t. \tag{5}$$

where I_x, I_{xx} are the partial derivatives of brightness in the x direction evaluated at (x, y, t). Rearranging terms we obtain an estimate for the brightness gradient in the x direction:

$$\hat{I}_x = \frac{I(x+\Delta x, y, t,)-I(x, y, t)}{\Delta x} = I_x + I_{xx} \Delta x + h.o.t. \tag{6}$$

The error $\epsilon_{I_x(sampling)}$ is defined as $\hat{I}_x - I_x$, the difference between the computed and true values. From (6), we obtain the approximate relationship

$$\epsilon_{I_x(sampling)} \approx I_{xx} \Delta x \ . \tag{7}$$

Likewise, the sampling error in the estimates of I_y and I_t are approximately given by

$$\epsilon_{I_y(sampling)} \approx I_{yy} \Delta y \tag{8}$$

$$\epsilon_{I_t(sampling)} \approx I_{tt} \Delta t \ . \tag{9}$$

The sampling error for the spatial gradients depends upon the spatial resolution of the camera, Δx and Δy, and the second spatial derivatives of the brightness function, I_{xx}, I_{yy}. The

sampling error for the temporal gradient, $e_{I_{t(\text{sampling})}}$, is influenced by the frame rate Δt and the higher order derivatives of the brightness function over time.

We can express $e_{I_{t(\text{sampling})}}$ purely in terms of spatial derivatives and motion. Differentiating the gradient constraint equation (3) with respect to x, y, and t we obtain the following three equations

$$I_{xx}u + I_x\frac{\partial u}{\partial x} + I_{yx}v + I_y\frac{\partial v}{\partial x} = -I_{tx} \tag{10}$$

$$I_{xy}u + I_x\frac{\partial u}{\partial y} + I_{yy}v + I_y\frac{\partial v}{\partial y} = -I_{ty} \tag{11}$$

$$I_{xt}u + I_x\frac{\partial u}{\partial t} + I_{yt}v + I_y\frac{\partial v}{\partial t} = -I_{tt} . \tag{12}$$

Where the second derivatives of the brightness function exist and are continuous, the left-hand sides of equations (10) and (11) can be substituted for I_{xt} and I_{yt} in (12). Collecting terms we see that

$$-\begin{bmatrix} u & v \end{bmatrix}\begin{bmatrix} I_{xx} & I_{xy} \\ I_{yx} & I_{yy} \end{bmatrix}\begin{bmatrix} u \\ v \end{bmatrix} - u\cdot\begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial v}{\partial x} \end{bmatrix}\begin{bmatrix} I_x \\ I_y \end{bmatrix} -$$

$$v\cdot\begin{bmatrix} \frac{\partial u}{\partial y} & \frac{\partial v}{\partial y} \end{bmatrix}\begin{bmatrix} I_x \\ I_y \end{bmatrix} + \begin{bmatrix} \frac{\partial u}{\partial t} & \frac{\partial v}{\partial t} \end{bmatrix}\begin{bmatrix} I_x \\ I_y \end{bmatrix} = -I_{tt} . \tag{13}$$

The first term in (13) depends upon optical flow while the rest of the left-hand side depends upon the derivatives of optical flow over time and space. If optical flow is approximately constant in a small neighborhood and approximately constant over time at each point on the image then

$$\begin{bmatrix} u & v \end{bmatrix}\begin{bmatrix} I_{xx} & I_{xy} \\ I_{yx} & I_{yy} \end{bmatrix}\begin{bmatrix} u \\ v \end{bmatrix} \approx I_{tt} . \tag{14}$$

Note the similarity between (6) and (14). We have derived a constraint equation for second derivatives that is analogous to gradient constraint equation.

Without loss of generality, we can rotate our coordinate system so that the flow vector at a point lies along the x-axis. In the new coordinate system we have

$$u^2 I_{xx} \approx I_{tt} . \tag{15}$$

It is evident from (15) that the magnitude of I_{tt} depends upon nonlinearities of the brightness function in the direction of motion and the magnitude of motion.

In summary, the systematic errors in the gradients which make up the coefficients of (4) are given by (7), (8), and (9). In general, the systematic error in estimating I_t is influenced by the magnitude of optical flow and the derivatives of optical flow and the first and second spatial derivatives of brightness. When the axis of coordinate system is aligned with motion and optical flow is nearly constant over time and space we can characterize the systematic error in the temporal derivative by

$$\epsilon_{I_t\text{compute}} \approx \nabla^2 I_{xx}\,\Delta x \tag{16}$$

If there is significant optical flow, the error given by (16) can become quite large in regions which contain nonlinearities in the brightness function and substantially alter our estimate of I_t.

4.2. Nonuniformity in the Flow Field.

The estimation scheme we have been analyzing assumes that velocity on the image plane is constant in some small neighborhood. This will be true only for very special surfaces and motions. When optical flow is not constant the method can provide a good approximation where flow varies slowly over small neighborhoods.

The true set of equations in (4) should actually be

$$\begin{aligned} I_x^{(i)}u \quad + \quad I_y^{(i)}v \quad &= -I_t^{(i)} \\ I_x^{(j)}(u+\Delta u) \; + \; I_y^{(j)}(v+\Delta v) \; &= -I_t^{(j)} \end{aligned} \tag{17}$$

where the actual flow vectors at points $\mathbf{p}_i$ and $\mathbf{p}_j$ are (u,v) and $(u+\Delta u, v+\Delta v)$, respectively, and the gradients are estimated at points $\mathbf{p}_i$ and $\mathbf{p}_j$. The difference between the true solution and our estimate can be treated as an error on the right-hand side of (17) by distributing the multiplication on the left-hand side of the system and rearranging terms as

$$\begin{aligned} I_x^{(i)}u \; + \; I_y^{(i)}v \; &= -I_t^{(i)} \\ I_x^{(j)}u \; + \; I_y^{(j)}v \; &= -I_t^{(j)} + \epsilon_{\Delta flow} \end{aligned} \tag{18}$$

where

$$\epsilon_{\Delta flow} = [I_x^{(j)}, I_y^{(j)}] \begin{bmatrix} \Delta u \\ \Delta v \end{bmatrix}. \tag{19}$$

Thus, the error caused by violation of the constant flow assumption can be treated as an additional error in the estimate of I_t.

To examine the significance of this error, we will consider size of $\epsilon_{\Delta flow}$ relative to $I_t^{(j)}$. But first we will convert to vector notation. Let

$$\mathbf{g}_i^{(j)} = \begin{bmatrix} I_x^{(j)} \\ I_y^{(j)} \end{bmatrix}, \quad \omega = \begin{bmatrix} u \\ v \end{bmatrix} \text{ and } \Delta\omega = \begin{bmatrix} \Delta u \\ \Delta v \end{bmatrix} \tag{20}$$

For the constraint equation at $\mathbf{p}_j$, we know from (17) and (19) that

$$\left| \frac{\epsilon_{\Delta flow}}{I_t^{(j)}} \right| = \frac{\left| \mathbf{g}_i^{(j)} \cdot \Delta\omega \right|}{\left| \mathbf{g}_i^{(j)} \cdot (\Delta\omega + \omega) \right|} \tag{21}$$

$$= \frac{\| \Delta\omega \| \cos\theta_1}{\| \Delta\omega + \omega \| \cos\theta_2} \tag{22}$$

where θ_1 is the angle between the gradient vector, $\mathbf{g}_i^{(j)}$, and the local change in optical flow,

$\Delta\omega$ and θ_e is the angle between the gradient vector, $\mathbf{g}_i$, and the vector $\Delta\omega_{xy}$.

The relative error in I_t depends upon the relative lengths of the vectors ω and $\Delta\omega$ and the degree to which each is magnified by the spatial gradient. In general the orientations of the spatial gradient, optical flow, and the local change in optical flow will be independent. So the spatial gradient will on the average magnify the flow vector in the same proportion as the change of flow vector. Therefore, on average, we expect the relative error in I_t to be strongly related to relative magnitudes of the flow and change of flow vectors.

In most scenes, flow will vary slowly over most of the image. At surface boundaries we can expect to frequently find discontinuities in optical flow due to discontinuities in motion or depth. Here, the variation in flow will contribute a substantial error and flow estimates will usually be quite poor. However, much of the image will consist of smoothly varying surfaces. When neighboring image points lie on the same smooth surface, flow will generally be similar and hence, the error contributed by variations in flow will be small.

We will consider an example which allows arbitrary three-dimensional translation of a planar surface to demonstrate the important factors influencing the error contributed by variations in optical flow. We consider two neighboring image points that lie on a surface translating with velocity (U,V,W) in three-dimensional space (see Figure A.1). Let the surface be defined by the planar equation

$$Z(X,Y) = R + \alpha X + \beta Y \ . \tag{23}$$

In appendix A we derive the following approximate bound

$$\frac{\| \Delta\omega \|}{\| \omega \|} \leq \tan\gamma \left(\| (\alpha,\beta) \| - \frac{| W |}{\| (U,V) \|} \right) \tag{24}$$

where,

$$\tan\gamma = \frac{\| (\Delta x, \Delta y) \|}{f} \ . \tag{25}$$

The angle γ is the angle subtended by $(\Delta x, \Delta y)$ with a focal length of f; this is simply the size of the neighborhood measured in degrees of visual angle. The length of the change-of-flow vector relative to the length of the flow vector depends upon the size of the neighborhood, the slope of the surface viewed, and the ratio of velocity along the line of sight to velocity perpendicular to the line of sight.

Recall that the value given by (24) represents a rough measure of the proportion of error on the right-hand side contributed by variations in optical flow. If the neighborhood is small we expect random errors in the temporal gradient to usually be larger than the error caused by flow variation. The gradient measurement errors discussed in the last section may lead to much larger degradation. So, for most of the image the error caused by variation in flow should not constitute a problem. However, at surface boundaries optical flow can change dramatically, especially when object motions are allowed. Here, the local optimization result will be a very poor measure of optical flow.

4.3. Ill-conditioning.

The accuracy of the estimates u and v will depend on the measurement errors in the gradient constraint equations and the error propagation characteristics of the linear system. When a system of linear equations is very sensitive to small errors in the coefficients or right-hand side it is said to be ill-conditioned. If the spatial intensity gradients change slowly then the linear system will contain constraint lines that are nearly parallel. As a consequence, the system will be nearly singular and small errors in the gradient measurements may result in large changes in the estimated flow value. We will find that the conditioning of the linear system largely depends upon nonlinearities in the brightness function which are perpendicular to the brightness gradient.

If the gradients are known exactly and optical flow is constant then

$$\mathbf{G}\omega = -\mathbf{b} \tag{26}$$

where,

$$\mathbf{G} = \begin{bmatrix} I_x^{(i)} & I_y^{(i)} \\ I_x^{(j)} & I_y^{(j)} \end{bmatrix}, \quad \omega = \begin{bmatrix} u \\ v \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} I_t^{(i)} \\ I_t^{(j)} \end{bmatrix}. \tag{27}$$

As before, the rows of $\mathbf{G}$ and $\mathbf{b}$ are taken from a point $\mathbf{p}_i$ and its neighbor $\mathbf{p}_j$. The vector ω will be in error to the degree that the gradient measurements are inaccurate and optical flow varies between points $\mathbf{p}_i$ and $\mathbf{p}_j$. The previous section showed that the error accrued when u and v are not constant is the same as that which would be obtained if the $\mathbf{b}$ vector is suitably modified as in (18). This error will be absorbed on the right-hand side of (26). Thus, the system which is actually solved is

$$(\mathbf{G} + \mathbf{E})(\omega + \delta\omega) = -(\mathbf{b} + \delta\mathbf{b}) \tag{28}$$

where,

$$\mathbf{E} = \begin{bmatrix} \epsilon_x^{(i)} & \epsilon_y^{(i)} \\ \epsilon_x^{(j)} & \epsilon_y^{(j)} \end{bmatrix}, \quad \delta\mathbf{b} = \begin{bmatrix} \epsilon_t^{(i)} \\ \epsilon_t^{(j)} \end{bmatrix} \text{ and } \delta\omega = \begin{bmatrix} \epsilon_u \\ \epsilon_v \end{bmatrix}. \tag{29}$$

The errors in the spatial and temporal gradients arise from both systematic and random measurement errors.

A number of measures of conditioning have been proposed [14]. The most widely used index of conditioning is the condition number, *cond*(), which is defined as

$$cond(\mathbf{G}) = \| \mathbf{G} \| \ \| \mathbf{G}^{-1} \| \tag{30}$$

for a matrix of coefficients $\mathbf{G}$. The condition number roughly estimates the extent to which relative errors in the coefficients and the righthand side are magnified in the estimate of optical flow. For the problem at hand, the conditioning of the matrix $\mathbf{G}$ is determined by the nature of the spatial brightness function over the interval $(\mathbf{p}_i, \mathbf{p}_j)$.

The inverse of $\mathbf{G}$ can be directed calculated as

$$\mathbf{G}^{-1} = \frac{1}{I_x^{(i)} I_y^{(j)} - I_x^{(j)} I_y^{(i)}} \begin{bmatrix} I_y^{(j)} & -I_x^{(j)} \\ -I_y^{(i)} & I_x^{(i)} \end{bmatrix} \tag{31}$$

$$= \frac{1}{\| \mathbf{g}^{(i)} \| \| \mathbf{g}^{(j)} \| \sin\phi} \begin{bmatrix} I_y^{(j)} & -I_x^{(j)} \\ -I_y^{(i)} & I_x^{(i)} \end{bmatrix} \tag{32}$$

where $\mathbf{g}^{(i)}$ is the spatial gradient vector at $\mathbf{p}_i$ and ϕ is the angle between $\mathbf{g}^{(i)}$ and $\mathbf{g}^{(j)}$.

Before we can evaluate the condition number we must select a matrix norm. We will use the Frobenius norm[1]. We will continue to use the $\| \cdot \|_2$ norm to evaluate vector norms. From the definition of $\| \cdot \|_F$ and the results above we have

$$cond(\mathbf{G}) = \frac{\| \mathbf{g}^{(i)} \|^2 + \| \mathbf{g}^{(j)} \|^2}{\| \mathbf{g}^{(i)} \| \| \mathbf{g}^{(j)} \| \sin\phi} \tag{33}$$

$$= \frac{1}{\sin\phi} \left(\frac{\| \mathbf{g}^{(i)} \|}{\| \mathbf{g}^{(j)} \|} + \frac{\| \mathbf{g}^{(j)} \|}{\| \mathbf{g}^{(i)} \|} \right) . \tag{34}$$

The magnitude of $cond(\mathbf{G})$ depends on the orientations and relative magnitudes of the two spatial gradient vectors. The value of $cond(\mathbf{G})$ is minimized when the spatial gradients are perpendicular and have the same magnitude. As the spatial gradients become more nearly parallel the magnitude of $cond(\mathbf{G})$ is increased, and hence, error propagation is worsened. Increases in the relative difference in the magnitudes of the spatial gradients also cause $cond(\mathbf{G})$ to increase. The magnitude of this effect will not usually be important. If the neither of the gradients is very small, then the relative sizes of the gradients will not differ enormously. The gradients will be poorly estimated where they are small, so for multiple reasons estimates will be error prone in these regions.

The most important factor determining conditioning is the angle between the gradients. Where the gradients are nearly parallel, conditioning will be a problem. Thus, if both points lie along a straight edge, we cannot obtain a solution. (This is an example of the aperture problem [11]).

Some higher derivatives of brightness must be large for there to be a significant change in gradient orientation over a small neighborhood. Let $\Delta\mathbf{g}$ be the difference between the two gradient vectors. We can expand the gradient in a Taylor series

$$\mathbf{g}^{(j)} = \mathbf{g}^{(i)} + \begin{bmatrix} I_{xx}^{(i)} & I_{yx}^{(i)} \\ I_{yx}^{(i)} & I_{yy}^{(i)} \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} + h.o.t. \tag{35}$$

Consequently,

$$\Delta\mathbf{g} \approx \begin{bmatrix} I_{xx}^{(i)} & I_{yx}^{(i)} \\ I_{yx}^{(i)} & I_{yy}^{(i)} \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} . \tag{36}$$

[1] The Frobenius norm, $\| \cdot \|_F$, is defined as the square root of the sum of the squares of all the elements. The Frobenius norm can be used to bound the more familiar $\| \cdot \|_2$ norm [15]. It can be shown that

$$\frac{1}{\sqrt{2}} \| \cdot \|_F \leq \| \cdot \|_2 \leq \| \cdot \|_F .$$

The angle between the gradients depends on the component of $\Delta \mathbf{g}$ that is perpendicular to $\mathbf{g}^{(1)}$. If optical flow is to be accurately estimated in a small region around $\mathbf{p}^{(1)}$, then at least one component of the second derivative perpendicular to the gradient must be large. There must be at least some direction in which we can select a neighbor so that the gradient orientation $\mathbf{g}^{(2)}$ will differ from $\mathbf{g}^{(1)}$.

4.4. Combining the Sources of Error

We now face a dilemma. We have just shown that some component of the second derivative must be large to minimize error propagation. However, we earlier showed that sampling errors in the gradients were proportional to the the magnitude of the second derivative. There is a tradeoff between the gradient measurement errors and conditioning. The problem would not be too serious if we were only concerned about errors in the spatial gradients. If we let the sampling interval be reasonably small with respect to the neighborhood from which we select our equations, we can potentially satisfy both goals -- the gradient can change slowly from pixel-to-pixel but the total variation over the neighborhood can be large enough to allow acceptable conditioning.

A serious conflict can arise in the tradeoff between conditioning and sampling errors in the temporal derivative. Recall that the systematic measurement error in $\hat{I}_t$ is proportional to nonlinearities in the spatial brightness function (13). To achieve acceptable conditioning, the spatial brightness function must be nonlinear in some direction. If optical flow is oriented in this direction, then the condition number and measurement errors will be inversely related. Increases in the magnitude of the second spatial derivatives will reduce the condition number and increase the measurement error. Note that there need not be a conflict; optical flow can be perpendicular to direction in which the gradient orientation is varying.

The problem is heightened by the sensitivity of measurements where the flow vector is large. The systematic measurement error in the temporal derivative increases as the square of flow magnitude (13). Where flow is large, even small nonlinearities can contribute significant measurement errors. However, where object points are stationary or moving slowly, the measurement error in the temporal gradient will be negligible and most accurate estimates will be obtained the gradients are not small and vary rapidly.

As an illustration of the interplay between the concerns of conditioning and measurement error, consider an image painted with an isotropic texture. If the region is stationary then a large amount of detail will be desirable to minimize conditioning. If optical flow is significantly greater than zero, then too much detail will lead to unacceptably large measurement errors. A balance must be struck between these two sources of error.

The conditioning of G can be improved by using a large neighborhood. The risk in choosing neighbors over too great a distance is that the error due to non-constant flow can become very large. If the neighbors lie on a single surface the contribution of errors due to non-constant flow will usually grow slowly with neighborhood size. But if neighbors lie on

different surfaces their motions may differ substantially. As neighborhood size is increased it becomes more likely that neighbors will lie across a surface boundary and the difference in optical flow will lead to significant errors.

The total error in the flow estimate is determined by the characteristics of the optical flow field, the nature of the brightness function, and the selection rule for constructing the linear system. The sources of error are summarized in Table 1.

Error Source	Determinants
1. Gradient Measurement Error	
(a) random (I_x, I_y, I_t)	(i) ↑↑ sensor noise
	(ii) ↑↑ quantization noise
(b) systematic (I_t)	(i) ↑↑ nonlinearities in the brightness function in the direction of optical flow
	(ii) ↑↑ optical flow magnitude
2. Non-constant Flow	(i) ↑↑ neighborhood size
	(ii) ↑↑ surface slant
	(iii) ↑↑ ratio of velocity along the line of sight to velocity perpendicular to the line of sight†
3. Ill-conditioning	(i) ↑↓ neighborhood size
	(ii) ↑↓ sin of the angle between the spatial gradient vectors
	(ii) ↑↑ relative difference in the magnitudes of the spatial gradient vectors

↑↑ error increases with determinant
↑↓ error decreases with determinant
† for translating surfaces

Table 1. The sources of error in local estimates of optical flow.

These factors interact in a complex way to determine the accuracy of the local optimization

scheme. Only where the contribution of these sources of error is balanced will good estimates be obtained.

5. Algorithm Extensions Based Upon the Error Analysis.

We next consider how knowledge about the causes of errors can be used to reduce errors and introduce techniques to judge the accuracy of estimates. The improvements in performance are based upon parameter selection and preprocessing of the image to extract the most information from a region while minimizing the intrusions of error. A method of iterative refinement [5, 16] is also described.

By examining the image sequence for the conditions which lead to errors we can judge the accuracy with which estimates can be made before the estimate is actually made. Examination of the flow estimate itself can provide additional information about the precision of the estimate. Together, *a priori* and *a posteriori* estimates of accuracy provide a useful heuristic for evaluating the precision of optical flow estimates.

5.1. Error Reduction Techniques

5.1.1. Smoothing

Blurring the image will reduce nonlinearities in the brightness function and consequently diminish the systematic error in the gradient estimates. Blurring will also worsen the propagation characteristics of the linear system causing random measurement errors and the errors due to non-constant flow to be magnified. Hence, blurring is desirable only in regions where the systematic error is predominant.

As noted in the last chapter, the systematic error in the gradients depends upon the nonlinearity of the brightness function over the sampling interval. For the temporal gradient, the systematic measurement error depends upon the linearity of the brightness function over the region which moves past a point of observation on the image and the variations of optical flow over time and space. Blurring will be most effective in portions of the image which undergo a significant motion and contain large nonlinearities in the brightness function. The degree of blurring should be sufficient to approximately linearize the brightness function over the region of translation.

The damage which blurring does to the conditioning of the linear system can be counterbalanced by increasing the size of the neighborhood over which the system is constructed. The risk incurred by enlarging the area from which the constraint equations are drawn is that the motions of the points may differ significantly, as could happen if points lied on two different surfaces. The selection of the radius of blur and the neighborhood size must be made judiciously so as to avoid increasing the error in the solution vector.

5.1.2. Over-determined Systems

Until this point we have ignored the problem of selecting the direction in which the neighbor is to be chosen to form the linear system. From our previous discussion of error propagation it is clear that the choice of direction can dramatically affect the error in the optical flow estimate. One way to circumvent the difficulty of choosing an appropriate direction is to construct an over-determined set of equations from points in many directions. The over-determined system can be solved by minimizing the residual over possible values of optical flow. The choice of the norm to be minimized and the minimization scheme may be an important determinant of the error, but are not analyzed here. As with two equation systems, conditioning will be important for over-determined systems and conditioning will be related to the same characteristics of the image as in the two equation case. Another approach is to perform the analysis separately in a number of directions and then seek a consensus among the solutions [17].

5.1.3. Iterative Registration

If optical flow is known approximately then this knowledge can be used to reduce the error in the local optimization technique. We develop a more general form of the gradient constraint equation that solves for the difference between an approximate estimate and the actual flow. Our derivation abbreviates an analysis presented by Paquin and Dubois [16].

Consider the image sequence that samples the three-dimensional image function — as pictured in Figure 1. We actually estimate the displacement of a point between successive samples of the image sequence. If velocity is constant then the displacement observed on the image over the time interval Δt is $(u\,\Delta t, v\,\Delta t)$. Let $\mathbf{d}$ be a displacement vector in 3-dimensional x, y, t-space. Let $\hat{\mathbf{d}}$ be an estimate of $\mathbf{d}$. Given a displacement estimate

$$\hat{\mathbf{d}} \equiv \begin{bmatrix} \hat{u}\,\Delta t \\ \hat{v}\,\Delta t \\ \Delta t \end{bmatrix} = \begin{bmatrix} x\text{-component of displacement} \\ y\text{-component of displacement} \\ t\text{-component of displacement} \end{bmatrix} \tag{37}$$

we can estimate optical flow by $(\hat{u}, \hat{v})$.

The vector $\dfrac{\hat{\mathbf{d}}}{\|\hat{\mathbf{d}}\|}$ is a unit vector in the direction of the estimated displacement. The gradient of I in this direction is

$$I_{\hat{d}} = \frac{\hat{\mathbf{d}}^T}{\|\hat{\mathbf{d}}\|}\begin{bmatrix} I_x \\ I_y \\ I_t \end{bmatrix} = \frac{1}{\|\hat{\mathbf{d}}\|}(I_x \hat{u} + I_y \hat{v} + I_t)\Delta t \tag{38}$$

$$= \frac{1}{\|\hat{\mathbf{d}}\|}(I_x \hat{u} + I_y \hat{v} - I_x u - I_y v)\Delta t \quad \text{(using (3))} \tag{39}$$

$$= \frac{1}{\|\hat{\mathbf{d}}\|}(I_x \delta u + I_y \delta v)\Delta t \tag{40}$$

where $\delta u = \hat{u} - u$ and $\delta v = \hat{v} - v$ are the errors in the estimated flow velocities. Finally, we get an expression that relates the error in the displacement estimate to measurable brightness gradients:

$$\|\hat{\mathbf{d}}\| I_{\hat{d}} = I_x \delta u\,\Delta t + I_y \delta v\,\Delta t = \begin{bmatrix} I_x \\ I_y \\ 0 \end{bmatrix} \cdot \Delta\hat{\mathbf{d}} \tag{41}$$

where

$$\Delta\mathbf{d} = \hat{\mathbf{d}} - \mathbf{d} = \begin{bmatrix} \delta\hat{u}\,\Delta t \\ \delta\hat{v}\,\Delta t \\ \Delta t \end{bmatrix}. \tag{42}$$

We can compute an estimate of the quantity (41) by using the Taylor expansion

$$I(x + \hat{u}\,\Delta t, y + \hat{v}\,\Delta t, t_1 + \Delta t) = I(x, y, t_1) + \|\hat{\mathbf{d}}\| I_{\hat{d}}. \tag{43}$$

Solving for I_d and combining with (41) yields the approximation

$$\begin{bmatrix} I_x \\ I_y \end{bmatrix} \cdot \begin{bmatrix} \delta x \\ \delta t \end{bmatrix} \Delta t = I(x + \hat{u}\,\Delta t, y + \hat{v}\,\Delta t, t_1 + \Delta t) - I(x, y, t_1) \tag{44}$$

The new constraint equation (41) is a more general form of the gradient constraint equation. The more general form relates the gradients in an arbitrary direction to the spatial gradients and optical flow. If the displacement estimate is $(0,0,\Delta t)$, then $I_d = I_t$.

We can use the general form of the gradient constraint equation to refine an estimate $\hat{\mathbf{d}}$ by solving (44) for $\Delta\mathbf{d}$. This process can be performed iteratively to find successively better estimates of optical flow. An improvement can be expected, on the average, whenever successive registrations are closer to the true displacement vector.

$$\| \Delta\mathbf{d}_{i+1} \| \leq \| \Delta\mathbf{d}_i \| \qquad i = 1,2,\ldots \tag{45}$$

The improvement arises from successively better estimates of I_d. As was demonstrated earlier in equation (13) the systematic error in the estimate of temporal derivative grows as the square of flow magnitude. The same relationship is true for direction derivative I_d and the flow difference in the general constraint equation.

Solving for the difference between an estimate of optical flow and the true optical flow is computationally equivalent to registering a portion of an image pair and estimating the change of position in the adjusted sequence. For this reason the technique has been called *iterative registration* [5]. The estimate of optical flow may be derived from estimates made at some previous time or from prior processing on a single frame pair.

Note that if the inequality of (45) does not hold, then the error might be expected to increase. If an estimate of optical flow is poor then the refinement effort may lead to an even larger error. The next section is devoted to methods to evaluate the quality of optical flow estimates. A measure of the accuracy of a flow estimate can be used to judge whether or not the estimate should be used for registration. Alternatively, the degree of registration can be base on the confidence put in the flow estimate, the more accurate the estimate is judged to be, the more that the frame pair should be adjusted in the direction of the estimate.

The iterative registration technique can be combined with variable blurring to produce a coarse-to-fine system for estimating optical flow [5]. Flow is roughly estimated with an image sequence which has been blurred so that the brightness function is approximately linear over areas the size of the maximum expected displacement. The coarse estimate of optical flow is used, at each point, to register a small region of the image at a finer level of resolution. This process is repeated at successively finer levels of resolution.

How much advantage can be gained from iterative registration? The spatial variation of optical flow will not be affected by registration. Thus, the error due to incompatibilities among equations in the linear system is unaffected by iterative registration. Also, the estimate of the directional gradient will contain some amount of random measurement error even if successive frames are in perfect registration. The propagation of these errors depends

primarily upon the conditioning of $|\mathbf{G}|$, which is not influenced by registration. We cannot expect to reduce the error in $\hat{\mathbf{d}}$ below that caused by random error in I_t and non-constant flow through iterative registration.

While preforming a coarse-to-fine registration the degree of blurring at each stage should be appropriate to the expected error in optical flow at the next more coarse level of analysis. In the absence of knowledge about the motions of individual points the blurring must be performed uniformly across the image. While the error will, on the average, be reduced for points which translate significantly, the error will tend to be increased for points which are stationary or move very little. No benefit is obtained by linearizing the brightness function at stationary regions and the error propagation characteristics are worsened. Some of the accuracy lost at stationary regions during coarse processing might be recovered at fine levels but, in general, the best estimates could be obtained at a fine level without registration. In the next section methods are developed to estimate the accuracy of optical flow estimates. This information can be used in the coarse-to-fine system of iterative registration to judge whether an improvement has been obtained at each level. A priori estimates of the magnitude of flow are also developed in the next section. The iterative registration technique can be improved by adapting the technique to knowledge about the accuracy of estimates and the magnitude of motion.

5.2. Estimating Error

Many of the factors which lead to errors in the local optimization estimation technique can be identified and measured from the image. The error propagation characteristics of the linear system can be estimated from the matrix of spatial gradients. The degree to which relative errors are magnified is indicated by $cond(\mathbf{G})$. Regions of the image for which the propagation characteristics are poor will be very sensitive to small measurement errors in the gradients. The optical flow estimates obtained in these regions are likely to be inaccurate.

The systematic measurement error in $\hat{I}_t$ was shown to depend upon the linearity of the brightness function in the direction of motion (13). One way to measure of the nonlinearity of the brightness function is to compare the spatial gradients of brightness in successive frames [2, 5]. If $I_x(x,y,t)$ is significantly different from $I_x(x,y,t+\delta t)$ then it can be inferred that the estimate of the temporal gradient is likely to be in error.

Once an estimate has been obtained we can bound the error by referring back to the image. The following *aposteriori* error bound can be derived from (44):

$$\| (\delta u, \delta v) \| \Delta t \geq \frac{I(x+\hat{u}\Delta t, y+\hat{v}\Delta t, t_1+\Delta t) - I(x,y,t_1)}{\| (I_x, I_y) \|} \tag{46}$$

If the norm of the spatial gradient is not too small, this will provide a good measure of the magnitude of the error in the flow estimate.

If an over-determined set of equations is used to estimate optical flow, then measurement errors in the gradients and incompatibilities among the constraint equations due to

differential motion will be reflected in the residual of the solution. The residual vector can be estimated by

$$\mathbf{G}\hat{\omega} - \mathbf{b} = \mathbf{r} \tag{47}$$

where $\hat{\omega}$ is the estimated optical flow and $\mathbf{r}$ is the residual. A large residual indicates that substantial errors exist in the system and that the estimated flow vector is likely to be inaccurate.

The residual vector will be especially large at occlusion edges where the change in flow is discontinuous. It has been proposed that the residual error be used as an indication of the presence of an occlusion edge [10]. To be identifiable, the change in optical flow across an occlusion edge must lead to an error which is greater than that normally encountered from other measurement errors. A threshold on the residual must be established which will normally be exceeded only at significant discontinuities in the flow field. The error accrued from a change in the flow vector is equivalent to a measurement error on the right-hand side of the local optimization system. Since the equivalent error on the right-hand side is magnified by the size of the spatial gradients, the threshold for identifying large residual errors should be adaptive to the spatial gradients. Likewise, it was shown that the systematic measurement errors in the gradients were related to the second derivatives of brightness, so the threshold on the residual should depend upon the second derivatives, as well.

5.3. Methods

The gradient-based approach is demonstrated with two versions of the local optimization technique. The basic local optimization method performs a least squares minimization on an over-determined set of gradient constraint equations to estimate optical flow at each point. Each image is first blurred with a gaussian blurring function. The standard deviation of the blurring function used to collect the data presented here was about 2 pixels. The blurring serves to reduce the noise in the image and linearize the brightness function.

Constraint equations from a group of neighboring points are gathered to produce an over-determined system of linear equations of the form

$$\mathbf{G}\omega = -\mathbf{b} \tag{48}$$

where,

$$\mathbf{G} = \begin{bmatrix} I_x & I_y \\ I_x & I_y \\ \vdots & \vdots \\ I_x & I_y \end{bmatrix}, \quad \omega = \begin{bmatrix} u \\ v \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} I_t \\ I_t \\ \vdots \\ I_t \end{bmatrix}. \tag{49}$$

Each row of $\mathbf{G}$ and $\mathbf{b}$, is evaluated at a different point. To insure that the equations are sufficiently distinct we selected neighbors from a 5×5 window centered around the point to be estimated.

In general, the over-determined system (48) has no exact solution. An approximate solution is found by minimizing the residual vector $\mathbf{r}$, defined in (47). The flow estimate is chosen to be the vector $\hat{\omega}$ which minimizes some criteria function of $\mathbf{r}$. In our work we minimize $\|\mathbf{r}\|_2$ by letting

$$\hat{\omega} = \mathbf{G}^{+}\mathbf{b} . \tag{50}$$

where $\mathbf{G}^{+}$ is the pseudo inverse of $\mathbf{G}$ [15]. Calculation of the pseudo inverse requires the inversion of the 2×2 matrix $\mathbf{G}^{t}\mathbf{G}$. The inverse will not exist where the local gradients do not sufficiently constrain optical flow to allow for an exact solution. In this case the confidence of the flow estimate is set to zero and u and v are undefined.

A confidence is assigned to each flow estimate on the basis of:

(a) an estimate of the measurement error in the temporal gradient,

(b) an estimate of the conditioning,

(c) the size of the residual vector $\mathbf{r}$, and

(d) the *a posteriori* bound given by (46)

The importance of each of these factors in determining the accuracy of estimates is discussed above. That analysis does not, however, provide us with a formula for estimating the total error in the flow vector (u,v). We must find a means to combine several factors which each indicate the presence of conditions which can be lead to errors.

Recall how each factor outlined above relates to the error in (u,v). The systematic *measurement error in the temporal gradient depends on the linearity of the brightness function*. The change in the spatial gradients between successive frames provides an indication of the linearity of the brightness function over the region which has translated by a point [5]. To obtain an estimate of the contribution of this error to errors in $\hat{\omega}$, we divide magnitude of the change in the spatial gradients by the magnitude of the spatial gradient.

The error propagation characteristics of the linear system $\mathbf{G}\hat{\omega} = \mathbf{b}$ can be determined by examining the matrix of spatial gradients. If linear system is ill-conditioned, small measurement errors will tend to produce large errors in $(\hat{u},\hat{v})$.

The residual vector indicates the degree to which the estimated flow vector jointly satisfies the system of constraint equations. But the value of the residual vector is not easy to interpret because the size of the residual is dependent of the overall magnitude of the brightness gradients. We normalize the residual by determining, for each equation, the minimum distance between the estimate and the equation. This is equal to the distance between the estimate and the constraint equation along a line perpendicular to the constraint equation that passes through the estimate. The average minimum distance is used as an index of the degree to which the equations are satisfied.

Once an estimate has been obtained, the *a posteriori* error bound given by (46) can be used to judge the accuracy of the estimate. In locations where this bound is large the computed optical flow vector is likely to be in error.

Each of the measurements described above provides an index of the expected error in the flow estimate. The four error estimates are not independent. The residual error and the *posteriori* bound measure the accumulative error from all sources in the flow estimate. The variation in the spatial gradient and the conditioning of G measure conditions which are likely to lead to poor estimates. Nonlinearity in the spatial brightness function is particularly troublesome for gradient measurement and the conditioning of G conveys the error propagation characteristics of the linear system. Even though the four estimates are not independent we found that they were best treated as separate sources of information and best combined multiplicatively. We examined a number of combination rules and found that the results were not highly sensitive to the particular rule for combining error estimates. A measure of confidence was obtained from the inverse of the error estimates. The confidence value can be interpreted as a rough measure of the likelihood that an optical flow estimate is correct.

5.3.1. Local Optimization with Iterative Registration

The simple method of local optimization can be extended by a method of iterative refinement. Flow estimates are used to register the frame pair on each successive iteration of the estimation procedure. It was earlier shown that the measurement error in the temporal gradient could be significantly reduced if the registration locally reduced the displacement of the image frames. Since the optical flow field will usually contain variations, the predicted registration will differ across the image. To obtain a consistent linear system, a small region of the first frame must be registered with the second frame on the basis of the predicted flow at the point for which optical flow is to be estimated. A system of linear equations is constructed from the registered region.

This process can be performed iteratively, using the optical flow estimates at the previous stage to register the frame pair on the next iteration. It is important to emphasize that, at each stage, the registration can only be expected to improve performance when the new registration is an improvement over the registration in the last iteration. Otherwise, the new estimate of optical flow will, in general, be worse then the previous estimate. Since it is desirable to register the image only where the flow estimates are believed to be correct, we register in proportion to the confidence in the flow estimate. A flow field of zero flow vectors is used to initialize the first iteration.

The iterative registration technique is employed with variable blurring to produce a coarse-to-fine system of analysis. Images are blurred with a gaussian weighting function. In early iterations the standard deviation of the gaussian weighting function is large. The standard deviation of the weighting function is reduced in each successive iteration. At each level, the radius of the blurring function should be large enough to guarantee that the brightness function is approximately linear over the maximum expected flow from the registered images.

The size of the neighborhood from which the constraint equations are selected must depend upon the amount which the images are blurred. At a coarse level of analysis there is

little detail which distinguishes nearby points. To obtain sufficiently different constraint equations the separation between observation points must be increased; otherwise, the conditioning of the linear system will degenerate.

Our system contains four iterations which correspond to four levels of coarseness. The neighbor size and the value of the standard deviation for the approximation to the gaussian weighting function are given in table 2 for each of the four iterations.

Iteration	Blur Radius σ	Neighborhood Size
1	7	6
2	5	4
3	3.5	3
4	2	2

Table 2 The coarse-to-fine analysis

A difficulty with the coarse-to-fine system is that the flow estimates for stationary and slowly moving points made at coarse levels may be worse than the initially assumed zero vector. To insure that the new flow estimate made at one level is not worse than the value input into the level, we examine the error bound given by (46) for both the initial and new estimates. If the error bound for the new estimate is significantly larger than the bound for the old estimate, it is ignored.

5.4. Results

The two methods described above were tested with the two image pairs presented in Figure 2. In the first sequence the camera was stationary. The two toy trains in the center of the first image move toward each other in the second image. The second sequence simulates a view from an aircraft flying over a city. The scene is actually a model of downtown Minneapolis. (This picture originally appeared in Barnard's thesis [18].)

The optical flow fields obtained with the simple local optimization technique are shown in Figure 3.a and Figure 3.b for the moving trains and flyover scenes. Associated with each vector is a confidence in the correctness of the value. A threshold on confidence was established which produced a reasonably dense sampling of mostly correct values. Only vectors which exceeded the confidence threshold are displayed. The resulting field was too dense to clearly display the entire field. Consequently, only 20% of the vector fields are shown in Figure 3.

The results of the coarse-to-fine method of iterative refinement are shown in Figures 3.c and 3.d. Confidence thresholds were established which produced vector densities which were comparable to that obtained with simple local optimization. Both techniques produce reasonably accurate results with the moving train sequence.

The two techniques are more easily distinguished on the basis of their performance with the flyover sequence. The simple local optimization method produces a large number of errors even for the relatively sparse sampling of vectors displayed in Figure 3.b. The method of iterative registration generated many fewer errors in fields which are much more dense than that obtained with the simple local optimization approach.

Note the areas where very few vectors are displayed. Optical flow is poorly estimated in these regions and low values of confidence are assigned to the estimates obtained there. The problematic regions are usually fit into one or more of the following characterizations

1. largely homogeneous regions,
2. highly textured regions which are moving, or
3. regions which contain large discontinuities in the flow field.

Optical flow estimates obtained in homogeneous areas are likely to be in error because of the poor conditioning of linear systems constructed in these regions. The temporal gradient is poorly measured in highly textured regions which undergo significant motion. In regions which contain large discontinuities in the flow field the temporal gradient is poorly estimated and the systems of equations from the region are likely to contain inconsistencies.

The success with which confidence estimates predict the accuracy of flow estimates is demonstrated in Figure 4. The flow field produced by the simple local optimization technique with the moving trains sequence is displayed in with a low threshold on confidence in Figure 4.a and a high threshold in Figure 4.b. As before, only 20% of the vectors which exceed the threshold are displayed. Similar thresholds are shown for the method of iterative registration in Figures 4.c and 4.d. For both methods confidence provides a reasonable index of the accuracy of flow estimates. A sparse sampling of accurate estimates exceeds the high confidence threshold. When the threshold is lowered, more dense fields are obtained with a significantly greater number bad vectors.

5.5. Summary.

The gradient constraint is a powerful tool for the analysis of dynamic imagery. Careful examination of one gradient-based technique led to a number of conclusions about the causes of errors, provided support for techniques to improve estimates, and indicated methods by which the accuracy of estimates could be judged. This analysis suggests that optical flow estimation should be adaptive to the nature of the brightness function and the characteristics of motion in a region of the image.

The results demonstrate the feasibility of measuring the quality of optical flow estimates. Gradient-based techniques are susceptible to a variety of problems and tend to produce very poor estimates in troublesome areas of the image. Without accurate estimates of confidence, good estimates can not be distinguished from bad and gradient-based techniques are of little use. This work emphasizes the importance of understanding the mechanisms which underlie computational methods. An awareness of the strengths and weaknesses of

methods and of the way in which they operate can lead to adaptations and enhancements which are of great practical value.

Appendix A. Optical Flow Variations

Several papers have examined the relationship between the three-dimensional motion of objects and observers and the characteristics of the optical flow field. We will consider an example which allows arbitrary three-dimensional translation of a planar surface to demonstrate the important factors influencing changes in optical flow over the image.

Let the three-dimensional coordinate system be attached to the camera as in Figure A.1 which is redrawn from Longuet-Higgins and Prazdny [19]. All motion is associated with the camera. Let U, V, and W be the translational velocities of the observer in the X, Y, and Z directions. When motion is constrained to translation, the components of the three-dimensional velocity vector are

$$X' = -U \qquad Y' = -V \qquad Z' = -W . \tag{A.1}$$

Using a perspective projection, the position of an object point on the image is related to its three-dimensional position by

$$x = \frac{fX}{Z} \qquad y = \frac{fY}{Z} \tag{A.2}$$

where f is the focal length of the camera. Velocity on the image plane, (u, v), at a point (x, y) is

$$u = x' \qquad v = y' . \tag{A.3}$$

Substituting from (A.2) into the right-hand side of (A.3) and differentiating we obtain

$$u = f \cdot \left(\frac{X'}{Z} - \frac{XZ'}{Z^2} \right) = \frac{-fU + xW}{Z} \tag{A.4}$$

and

$$v = f \cdot \left(\frac{Y'}{Z} - \frac{YZ'}{Z^2} \right) = \frac{-fV + yW}{Z} \tag{A.5}$$

Consider a point $\mathbf{P}_0$ on the surface of a rigid body which projects to $\mathbf{p}_0$ on the image. We orient the coordinate system so that $\mathbf{P}_0$ lies on the observer's line of sight. The three-

This coordinate transformation is not strictly correct for a planar retina as pictured in Figure A.1. The change of coordinates can be justified in several ways. It can be assumed that the retina is globally spherical but can locally be modeled as planar. Or, it can be assumed that the distance of $\mathbf{p}_0$ from the origin is sufficiently small relative to the focal length that the distortion introduced by the transform will be minimal. Or finally, we can simply restrict our attention to points along the line of sight.

dimensional coordinates of P_0 are $(0,0,R)$ and the position of p_0 on the image is $(0,0)$. We assume that the surface is planar so that

$$Z(X,Y) = R - \alpha X - \beta Y \tag{A.6}$$

for points on the surface near P_0.

Following Longuet-Higgins and Prazdny, we introduce the dimensionless coordinate

$$z = f \cdot \left(\frac{Z - R}{Z} \right) = \alpha x + \beta y \ . \tag{A.7}$$

The components of optical flow formalized in (A.4) and (A.5) can be rewritten as

$$u = \left(\frac{-fU + xW}{R} \right) \left(1 - \frac{z}{f} \right) \tag{A.8}$$

and

$$v = \left(\frac{-fV + yW}{R} \right) \left(1 - \frac{z}{f} \right) . \tag{A.9}$$

The surface is assumed to be planar, so the derivatives of u and v with respect to x and y are well defined. At the point p_0, where $x = y = z = 0$, u and v are

$$u = -\frac{fU}{R} \quad \text{and} \quad v = -\frac{fV}{R} . \tag{A.10}$$

The derivatives of u and v are given by

$$u_x = \frac{\alpha U + W}{R} \qquad u_y = \frac{\beta U}{R} \tag{A.11}$$

$$v_x = \frac{\alpha V}{R} \quad \text{and,} \quad v_y = \frac{\beta V + W}{R} \tag{A.12}$$

since

$$z_x = \alpha \quad \text{and} \quad z_y = \beta. \tag{A.13}$$

Recall that the error incurred by assuming constant flow could be treated as measurement error in I_t, on the right-hand side of (18). The magnitude of this error, relative to I_t, is strongly dependent on the ratio of the magnitude of the change in optical flow to the magnitude of the flow vector. We can now express the ratio of change-of-flow to flow in terms of the three-dimensional parameters of shape and motion, and the viewing angle. The change in optical flow between two points separated by $(\Delta x, \Delta y)$ is

$$(\Delta u, \Delta v) = (\Delta x u_x + \Delta y u_y, \Delta x v_x + \Delta y v_y) . \tag{A.14}$$

Inserting the appropriate terms from (A.11) and (A.12) into (A.14) and dividing by optical flow as given by (A.10), we arrive at an expression for the ratio of change-of-flow to flow at a point:

$$\frac{\left\| \begin{bmatrix} \Delta u \\ \Delta v \end{bmatrix} \right\|}{\left\| \begin{bmatrix} u \\ v \end{bmatrix} \right\|} = \frac{\left\| \begin{bmatrix} (\alpha U + W)\Delta x + \beta U \Delta y \\ \alpha V \Delta x + (\beta V + W)\Delta y \end{bmatrix} \right\|}{\left\| \begin{bmatrix} fU \\ fV \end{bmatrix} \right\|} \tag{A.15}$$

$$= \frac{\left\| \begin{bmatrix} U \\ V \end{bmatrix} \left| \alpha,\beta \right| \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} \right\| - W \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} \Bigg\|}{f \left\| \begin{bmatrix} U \\ V \end{bmatrix} \right\|} \tag{A.16}$$

$$\leq \tan\gamma \left(\| \left| \alpha,\beta \right| \| + \frac{|W|}{\left\| \begin{bmatrix} U \\ V \end{bmatrix} \right\|} \right) \tag{A.17}$$

where,

$$\tan\gamma = \frac{\left\| \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} \right\|}{f} . \tag{A.18}$$

The angle γ is the angle subtended by $(\Delta x, \Delta y)$ with a focal length of f ; this is simply the size of the neighborhood measured in degrees of visual angle. The length of the change-of-flow vector relative to the length of the flow vector depends upon the size of the neighborhood, the slope of the surface viewed, and the ratio of velocity along the line of sight to velocity perpendicular to the line of sight.

References

1. C. Cafforio and F. Rocca, "Methods for measuring small displacements of television images," *IEEE Trans. Information Theory*, vol. IT-22, pp. 573-579, September 1976.
2. C.L. Fennema and W.B. Thompson, "Velocity determination in scenes containing several moving objects," *Computer Graphics and Image Processing*, vol. 9, pp. 301-315, April 1979.
3. B.K.P. Horn and B. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, pp. 185-203, 1981.
4. J.O. Limb and J.A. Murphy, "Estimating the velocity of moving images in television signals," *Computer Graphics and Image Processing*, vol. 4, pp. 311-327, December 1975.
5. B. D. Lucas and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision," *Proceedings of the 5th International Joint Conference on Artificial Intelligence*, pp. 674-679, August, 1981.
6. A. N. Netravali and J. D. Robbins, "Motion-compensated television coding: part I," *The Bell System Technical Journal*, vol. 58, no. 3, March, 1979.
7. R.J. Schalkoff, *Algorithms for a real-time automatic video tracking system*, May 1979. Ph.D. Dissertation, University of Virginia
8. R. J. Schalkoff and E. S. McVey, "A Model and Tracking Algorithm for a Class of Video Targets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI 1-4, no. 1, pp. 2-10, January, 1982.
9. B. G. Schunck and B.K.P. Horn, "Constraints on optical flow," *Proceedings IEEE conference on Pattern Recognition and Image Processing*, pp. 205-210, Aug. 1981.
10. W.B. Thompson and S.T. Barnard, "Low-level estimation and interpretation of visual motion," *Computer*, August 1981.

11 Ellen C Hildreth "Computations Underlying the Measurement of Visual Motion." *Artificial Intelligence* vol 23, pp 309-354 1984

12 Joseph K Kearney and William B Thompson, "Gradient-based Estimation of Optical Flow with Global Optimization." *The First Conference on Artificial Intelligence Applications*, pp. 376-380. Denver, CO, Dec 1984

13 J.K Kearney, *Gradient-based Estimation of Optical Flow*, University of Minnesota August, 1983. Ph.D. dissertation

14 Joan R. Westlake, *A Handbook of Numerical Matrix Inversion and Solution of Linear Equations*, John Wiley and Sons, Inc., New York, NY, 1968.

15. G.W. Stewart, *Introduction to Matrix Computations*, Acedemic Press, Inc., New York, NY, 1973

16. R. Paquin and E. Dubois, "A spatio-temporal gradient method for estimating the displacement field in time-varying imagery," *Computer Vision, Graphics, and Image Processing*, vol 21, no. 2, pp. 205-221, Feb 1983.

17. M. A. Fischler and R. C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," *Communications of the ACM*, vol. 24, pp. 381-395, June 1981.

18 S.T. Barnard, *The image correspondence problem*, Computer Science Department, University of Minnesota, December 1979. Ph.D. dissertation

19. H.C. Longuet-Higgins and K. Prazdny, "The interpretation of a moving retinal image," *Proc. R. Soc. Lond.* vol. B 208, pp. 385-397, 1980.

Figure Captions

Figure 1. The sampled image function.

Figure 2. Image Sequences.

Figure 3. Optical flow estimates.

Figure 4. The accuracy of confidence estimates. Optical flow estimates exceeding low and high thresholds on confidence are displayed.

Figure A.1 The camera-based coordinate system.

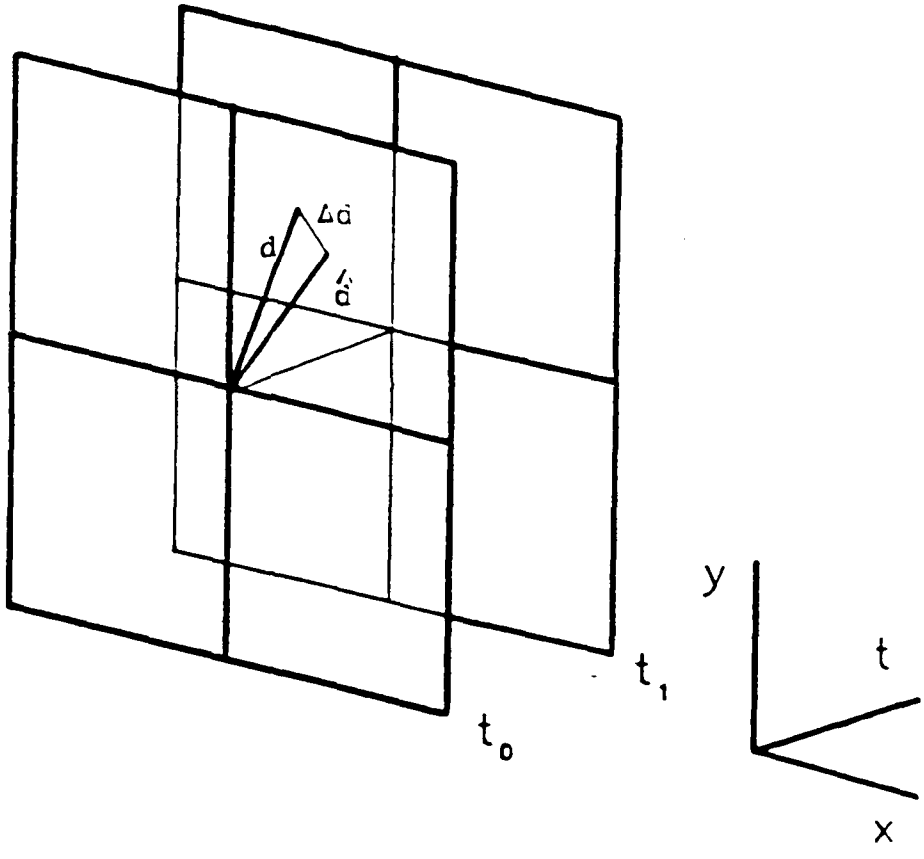
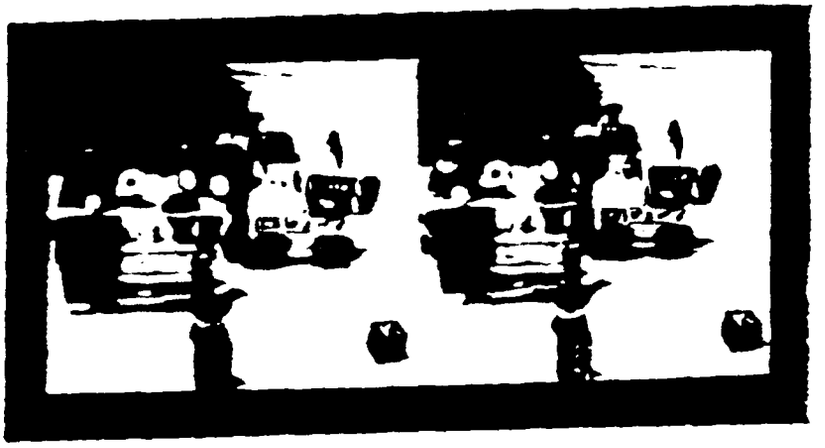


Figure 1.



(a) moving trains

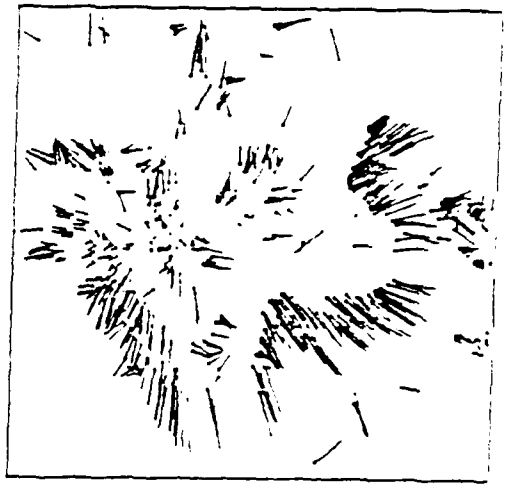
(b) simulated flyover

Figure 2.

(a) simple local optimization

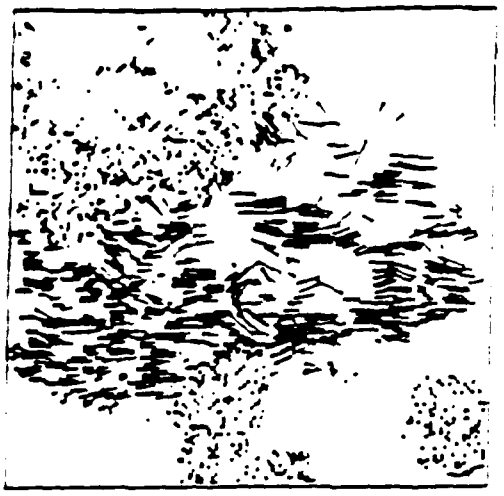
(b) simple local optimization

(c) iterative registration

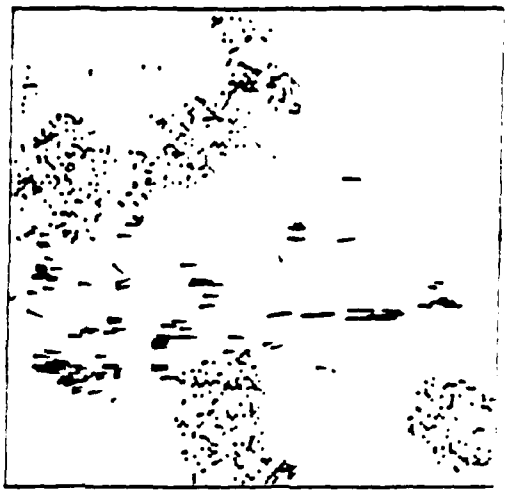


(d) iterative registration

Figure 3.

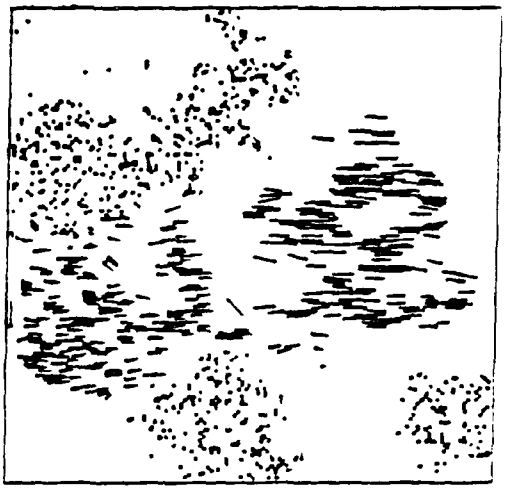


(a) simple local optimization
(low threshold)



(b) simple local optimization
(high threshold)

(c) iterative registration
(low threshold)



(d) iterative registration
(high threshold)

Figure 4.

Dynamic Occlusion Analysis in Optical Flow Fields

WILLIAM B. THOMPSON, MEMBER, IEEE, KATHLEEN M. MUTCH, MEMBER, IEEE, AND VALDIS A. BERZINS, MEMBER, IEEE

Abstract—Optical flow can be used to locate dynamic occlusion boundaries in an image sequence. We derive an edge detection algorithm sensitive to changes in flow fields likely to be associated with occlusion. The algorithm is patterned after the Marr-Hildreth zero-crossing detectors currently used to locate boundaries in scalar fields. Zero-crossing detectors are extended to identify changes in direction and/or magnitude in a vector-valued flow field. As a result, the detector works for flow boundaries generated due to the relative motion of two overlapping surfaces, as well as the simpler case of motion parallax due to a sensor moving through an otherwise stationary environment. We then show how the approach can be extended to identify which side of a dynamic occlusion boundary corresponds to the occluding surface. The fundamental principal involved is that at an occlusion boundary, the image of the surface boundary moves with the image of the occluding surface. Such information is important in interpreting dynamic scenes. Results are demonstrated on optical flow fields automatically computed from real image sequences.

Index Terms—Dynamic occlusion, dynamic scene analysis, edge detection, optical flow, visual motion.

I. Introduction

An optical flow field specifies the velocity of the image of points on a sensor plane due to the motion of the sensor and/or visible objects. Optical flow can be used to estimate aspects of sensor and object motion, the position and orientation of visible surfaces relative to the sensor, and the relative position of different objects in the field of view. As a result, the determination and analysis of optical flow is an important part of dynamic image analysis. In this paper, we develop an operator for finding occlusion boundaries in optical flow fields. We deal exclusively with *dynamic occlusions* in which flow properties differ on either side of the boundary. The operator is effective for both motion parallax, when a sensor is moving through an otherwise stationary environment, and for more general motion in which multiple moving objects can be in the field of view. The multiple moving object situation is more difficult because boundaries are marked by almost arbitrary combinations of changes in magnitude and/or direction of flow. The technique is extended so that a determination may be made about which side of a dynamic occlusion boundary corresponds to the occluding surface. Such a determination is of great importance for interpreting the shape and spatial organization of visible surfaces. Results are demonstrated on real image sequences with flow fields computed using the token matching technique described in [1]. Reliability is obtained by dealing only with methods able to integrate flow field information over relatively large neighborhoods so as to reduce the intrinsic noise in fields determined from real image sequences.

Manuscript received June 1, 1984; revised February 14, 1985. Recommended for acceptance by W. E. L. Grimson. This work was supported by the National Science Foundation under Grant MCS-81-05215, the Air Force Office of Scientific Research under Contract F49620-83-0140, and by Zonta International.

W. B. Thompson and V. A. Berzins are with the Department of Computer Science, University of Minnesota, Minneapolis, MN 55455.

K. M. Mutch was with the Department of Computer Science, University of Minnesota, Minneapolis, MN 55455. She is now with the Department of Computer Science, Arizona State University, Tempe, AZ 85287.

II. Boundary Detection

Conventional edge operators detect discontinuities in image luminence. These discontinuities are difficult to interpret, however, because of the large number of factors that can produce luminence changes. Boundaries in optical flow can arise from many fewer causes and, hence, are often more informative. If a sensor is moving through an otherwise static scene, a discontinuity in optical flow occurs only if there is a discontinuity in the distance from the sensor to the visible surfaces on either side of the flow boundary [2]. Discontinuities in flow will occur for all visible discontinuities in depth, except for viewing angles directly toward or away from the direction of sensor motion. If objects are moving with respect to one another in the scene, then all discontinuities in optical flow correspond either to depth discontinuities or surface boundaries, and most depth discontinuities correspond to flow discontinuities.

The use of local operators to detect discontinuities in optical flow has been suggested by others. Nakayama and Loomis [3] propose a "convexity function" to detect discontinuities in image plane velocities generated by a moving observer. Their function is a local operator with a center-surround form. That is, the velocity integrated over a band surrounding the center of the region is subtracted from the velocity integrated over the center. The specifics of the operator are not precisely stated, but a claim is made [3, Fig. 3] that the operator returns a positive value at flow discontinuities. (In fact, most reasonable formulations of their operator would yield a value of 0 at the boundary, with a positive value to one side or the other.) Clocksin [2] develops an analysis of optical flow fields generated when an observer translates in a static environment. He shows that, in such circumstances, discontinuities in the magnitude of flow can be detected with a Laplacian operator. In particular, singularities in the Laplacian occur at discontinuities in the flow. He also showed that, in this restricted environ-

0162-8828/85/0700-0374$01.00

ment, the magnitude of optical flow at a particular image point is inversely proportional to distance, and the distances can be recovered to within a scale factor of observer speed. It is thus trivial to determine which of two surfaces at an edge is occluding, for example, by simply comparing magnitudes of the two surface velocities, even when observer speed is unknown.

For this restricted situation in which a sensor moves through an otherwise static world

$$flow(\vec{x}) = f_r(\vec{x}) + \frac{f_t(\vec{x})}{r(\vec{x})} \tag{1}$$

where at an image point $\vec{x}$, $flow(\vec{x})$ is the optical flow (a two-dimensional vector), f_r is the component of the flow due to the rotation of the scene with respect to the sensor, f_t is dependent on the translational motion of the sensor and the viewing angle relative to the direction of translation, and r is the distance between the sensor and the surface visible at $\vec{x}$ [4]. For a fixed $\vec{x}$, flow varies inversely with distance. Both f_r and f_t vary slowly (and continuously) with $\vec{x}$. Discontinuities in *flow* thus correspond to discontinuities in r. Furthermore, it is sufficient to look only for discontinuities in the magnitude of *flow*. This relationship holds only for relative motion between the sensor and a single, rigid structure. When multiple moving objects are present, (1) must be modified so that there is a separate $f_r^{(i)}$ and $f_t^{(i)}$ specifying the relative motion between the sensor and each rigid object. Discontinuities associated with object boundaries may now be manifested in the magnitude and/or direction of *flow*.

Boundary detectors for optical flow fields should satisfy two criteria: 1) sensitivity to rapid spatial change in one or both of the magnitude and direction of flow, and 2) operation over a sufficiently large neighborhood to reduce sensitivity to noise in computed flow fields. It is desirable to achieve the second criterion without an unnecessary loss of spatial resolution in locating the boundary or a need for postprocessing to reduce the width of detected boundaries. The *zero-crossing* detectors of Marr and Hildreth [5] may be extended to optical flow fields in a manner that achieves both objectives [6]. For scalar fields (e.g., intensity images), zero-crossing edge detection proceeds as follows. 1) Smooth the field using a symmetrical Gaussian kernel. 2) Compute the Laplacian of the smoothed function. 3) Look for directional zero crossings of the resulting function (e.g., look for points at which, along some direction, the function changes sign). Under a set of relatively weak assumptions, these zero crossings can be shown to correspond to points of most rapid change in some direction in the original function. The convolution with a Gaussian provides substantial noise reduction and, in addition, allows tuning of the method for edges of a particular scale. Steps 1) and 2) involve evaluating the function $\nabla^2 G * I$, where G is a Gaussian kernel, $*$ is the convolution operation, and I is the original image. The effect of the $\nabla^2 G$ operator can be approximated by blurring the original function with two different Gaussian kernels of appropriate standard deviation, and then taking the difference of the result. This formulation results in computational simplifications [7], [8] and also corresponds nicely to several physiological models that have been proposed for early visual processing.

The effect of this approach is to identify edge points where the intensity of the blurred image is locally steepest. More precisely, an edge can be defined as a peak in the first directional derivative, or as a zero crossing in the second directional derivative. At an edge, the second directional derivative has zero crossings in almost all directions, but the preferred direction is normal to the locus of the zero crossings, which is the same as the direction where the zero crossing is steepest for linearly varying fields [5]. For vector images such as optical flow fields, the directional derivatives are vector valued, and we want the *magnitude* of the first directional derivative to have a peak.

This extension to two-dimensional flow fields is relatively straightforward. The optical flow field is first split into separate scalar components corresponding to motion in the x and y directions. The $\nabla^2 G$ operator is applied to each of these component images, and the results combined into a componentwise Laplacian of the original flow field. (The Laplacian is a vector operator which can be expressed in arbitrary coordinate systems. For convenience, we choose a Cartesian coordinate system.) This componentwise Laplacian operation is implemented by subtracting two componentwise blurred versions of the original. With the proper set of weak assumptions, discontinuities in optical flow correspond to zeros in both of these component Laplacian fields. At least one of the components will have an actual zero crossing. The other will have either a zero crossing or will have a constant zero value in a neighborhood of the discontinuity. If the componentwise Laplacians are treated as a two-dimensional vector field, discontinuities are indicated by directional reversals in the combined field. Because of the discrete spatial sampling and a variety of noise sources, the zeros or zero crossings in the two components of the field may not actually be exactly spatially coincident. Thus, exact reversal is not expected, and a range of direction changes of about 180° is accepted. A threshold on the sum of the vector magnitudes at the location of the flip is used to ensure that the zero crossing is of significant slope. This is analogous to the threshold on zero-crossing slope which is often used in practice when zero-crossing techniques are used on intensity images, and serves to filter out small discontinuities.

The approximations made by the computations described above will be good if the variation of the field parallel to the edge is much more uniform than the variation normal to the edge. For scalar images, exact results will be obtained if the intensity varies at most linearly along the edge contour [5]. For vector images, the field must vary at most linearly in some *neighborhood* of the edge contour, so that the assumptions required are slightly stronger than for scalar images. Appendix I contains the analysis for the case of vector images.

Two examples of this technique applied to real images are shown below. In both examples, the objects are toy animals with flat surfaces, shown moving in front of a textured background. In Fig. 1(a), the tiger translates parallel to the image plane from right to left between frames 1 and 2. The elephant rises off its front legs between frames 1 and 2, effectively rotating about an axis at its hind feet oriented perpendicularly to the image plane. The elephant also translates slightly to the left parallel to the image plane. The optical flow vectors, shown

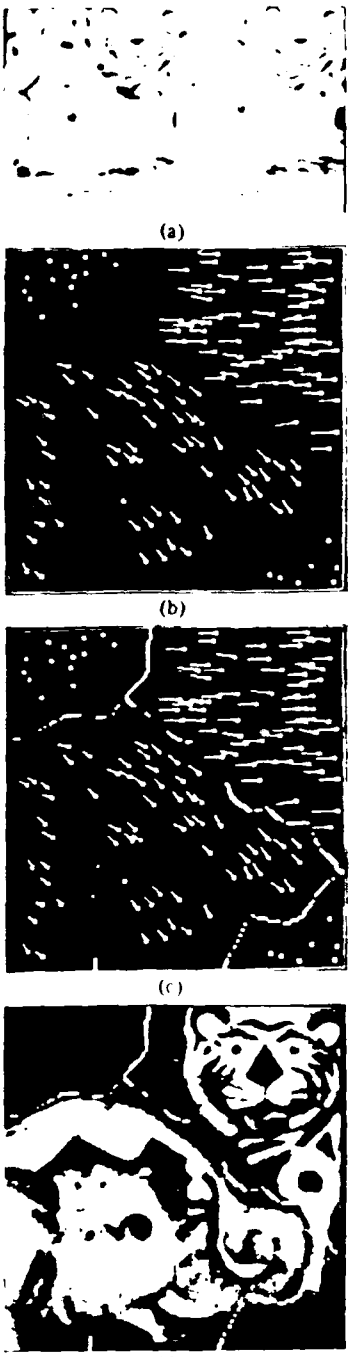


Fig. 1. (a) Image pair. (b) Optical flow. (c) Detected edge overlaid onto flow field. (d) Dectected edge overlaid onto first frame of sequence.

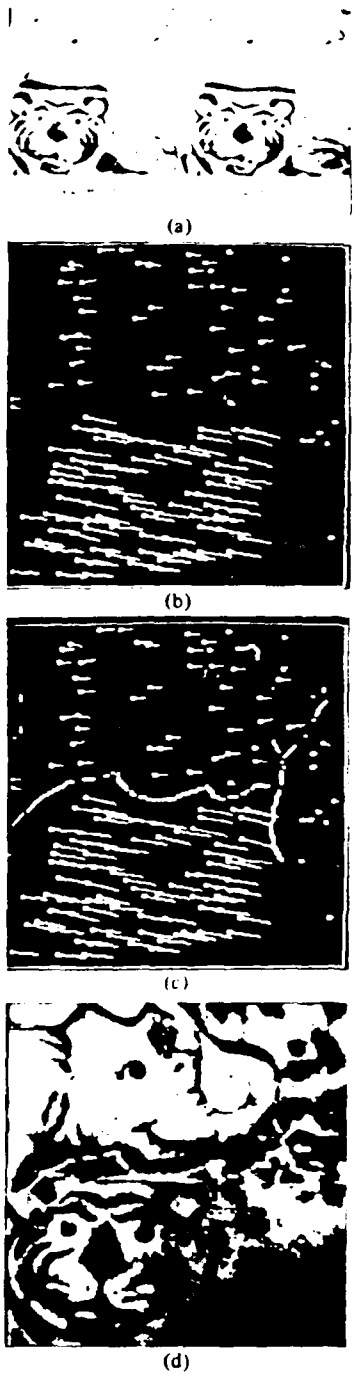


Fig. 2. (a) Image pair. (b) Optical flow. (c) Detected edge overlaid onto flow field. (d) Detected edge overlaid onto first frame of sequence.

in Fig. 1(b), were obtained by relaxation labeling token matching, as described in [1]. Notice that the flow vectors on the elephant and tiger have approximately the same magnitude but differ in direction. Each component of this flow field was convolved with approximated Gaussians of standard deviations 3.65 and 5.77. The ratio of these standard deviations is 1 : 1.6. The two convolved flow fields were subtracted, and the resulting vector field was searched for reversals in vector direction. A boundary strength threshold was chosen to eliminate noise points due to small, local variations in estimated flow. In Fig. 1(c), the points where reversals were found are shown overlaid on the original flow field, and in Fig. 1(d) the points are overlaid in white on the first image of the pair. The edge points form a good boundary between the discontinuous optical flow vector fields [Fig. 1(c)]; but because these fields are so sparse, the edge points match only the approximate locations of the true edges [Fig. 1(d)].

In Fig. 2(a), both the tiger and elephant are translating to the right, parallel to the image plane between frames 1 and 2. The flow field shown in Fig. 2(b) was obtained in the same manner as in Fig. 1(b). The direction of the flow vectors on both animals is approximately the same, but there is a dis-

continuity in magnitude. Two Gaussian filtered versions of the flow fields were obtained with standard deviations of 3.16 and 5.16, a ratio of 1:1.6. The locations of vector reversals resulting from differencing the two filtered fields are shown in Fig. 2(c) and (d).

The width of the Gaussian kernel used in the $\nabla^2 G$ operator, the density of the computed optical flow field, and the spatial variability of flow all interact to affect the performance of the boundary detection. As with the use of zero-crossing detectors for scalar fields, it may be desirable to use a range of kernel sizes and then combine the results to obtain a more robust indicator for the presence of a boundary. While zero-crossing contours are, in principle, connected, the use of a threshold on the slope at the zero crossing results in some portions of the boundary being missed. In practice, zero-crossing boundary detection for both scalar and vector fields often requires such thresholds to avoid significant problems with false boundary indications in slowly varying regions of the fields. Work still needs to be done on better techniques for selecting zero crossings that correspond to true boundaries.

III. Identifying Occluding Surfaces

When analyzing edges between dissimilar image regions that arise due to occlusion boundaries, it is important to determine which side of the edge corresponds to the occluding surface. Occlusion boundaries arise due to geometric properties of the occluding surface, not the occluded surface. Thus, while the shape of the edge provides significant information on the structure of the occluding surface, it says little or nothing about the structure of the surface being occluded. In situations where a sensor is translating through an otherwise static scene, any significant local decrease in r in (1) increases the magnitude of *flow*. Thus, at a flow boundary, the side having the larger magnitude of flow will be closer, and thus will be occluding the farther surface. Sensor rotation complicates the analysis, while if objects in the field of view move with respect to each other, there is no direct relationship between magnitude of *flow* and r. Surfaces corresponding to regions on opposite sides of a boundary may move in arbitrary and unrelated ways. However, by considering the flow values on either side of the boundary and the manner in which the boundary itself changes over time, it is usually possible to find which side of the boundary corresponds to the occluding surface, although the depth to the surfaces on either side cannot be determined.

The principle underlying the approach is that the image of the occluding contour moves with the image of the occluding surface. Fig. 3 illustrates the effect for simple translational motion. Shown on the figure are the optical flow of points on each surface and the flow of points on the image of the boundary. In Fig. 3(a), the left surface is in front and occluding the surface to the right. In Fig. 3(b), although the flow values associated with each surface are the same, the left surface is now behind and being occluded by the surface to the right. The occluding surface cannot be determined using only the flow in the immediate vicinity of the boundary. The two cases can be distinguished because, in Fig. 3(a), the flow boundary determined by the next pair of images will be displaced to the left, while in Fig. 3(b) it will be displaced to the right.

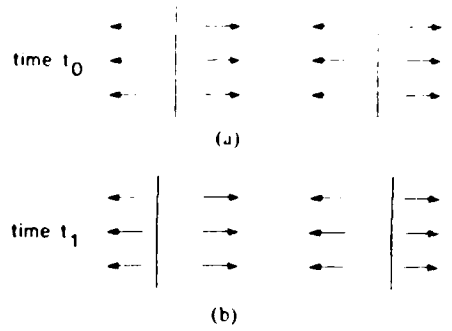


Fig. 3. Optical flow at a boundary at two instants in time. (a) Surface to the left is in front. (b) Surface to the right is in front.

To formalize the analysis, we need to distinguish the optical flow of the boundary itself from the optical flow of surface points. The flow of the boundary is the image plane motion of the boundary, which need not have any direct relationship to the optical flow of regions adjacent to the boundary. The magnitude of the optical flow of boundary points parallel to the direction of the boundary typically cannot be determined, particularly for linear sections of boundary. Thus, we will limit the analysis in this section to the component of optical flow perpendicular to the direction of the image of occlusion boundaries. As a result, if the flow on both sides of the boundary is parallel to the boundary, the boundary will still be detectable, but the method given here will provide no useful information about which surface is occluding.

We can now state the basic principle more precisely. Choose a coordinate system in the image plane with the origin at a particular boundary point and the x axis oriented normal to the boundary contour, with $x > 0$ for the occluding surface. The camera points in the z direction, and the image plane is at $z = 0$. Let $f_x(x, y)$ be the x component of optical flow at the point (x, y). Let f_b be the x component of the flow of the boundary itself at the origin (i.e., f_b is the image plane velocity of the boundary in a direction perpendicular to the boundary). Then, for rigid objects,

$$f_b = \lim_{x \to 0+} f_x(x, 0) = f_x(0, 0). \tag{2}$$

We will show that this relationship is true for arbitrary rigid body motion under an orthographic projection. For a single smooth surface, perspective projections are locally essentially equivalent to a rotation plus a scale change, although the analysis is more complex. Equation (2) specifies a purely local constraint and, as the limit is taken from only one side of the boundary, is dependent on flow values on a single surface. Thus, the limit result will hold as well for perspective projections. Algorithms which utilize the result in (2) may suffer, however, if properties of more than a truly local area of the field are utilized. The instantaneous motion of a rigid object relative to a fixed coordinate system can be described with respect to a six-dimensional, orthogonal basis set. Three values specify translational velocity, the other three specify angular velocity. These six coordinates of motion can be conveniently classified into four types: translation at constant depth, translation in depth, rotation at constant depth, and rotation in depth. Translation at constant depth is translation in a direction parallel to the image plane. Translation in depth is translation per-

pendicular to the image plane. Rotation at constant depth is rotation around an axis perpendicular to the image plane. Rotation in depth is rotation around an axis parallel to the image plane. Any instantaneous motion can be described as a combination of these four types. For orthographic projections, translation in depth has no effect on the image. Thus, we need to show that the above relationship relating boundary and surface flow holds for the three remaining motion types.

A point on the surface of an object in the scene that projects into a boundary point in the image will be referred to as a *generating point* of the occlusion boundary. The family of generating points defines a *generating contour*, which lies along the extremal boundary of the object with respect to the sensor. For both translation and rotation at constant depth, the generating contour remains fixed to the occluding surface over time. Thus, the boundary and adjacent points move with exactly the same motion. As a result, the projection of the surface flow in the direction normal to a particular boundary point is identical to the projection of the boundary flow in the same direction. (The result is strictly true only for instantaneous flow. Over discrete time steps, boundary curvature will affect the projected displacement of the boundary.)

The analysis of rotation in depth is complicated by a need to distinguish between sharp and smooth occlusion boundaries, based on the curvature of the occluding surface. The intersection of the surface of the object and a plane passing through the line of sight to the generating point and the surface normal at the generating point defines a *cross section contour*. The cross section contour and the generating contour cross at right angles at the generating point. *Sharp* boundaries occur when the curvature of the cross section contour at a generating point is infinite. *Smooth* boundaries occur when the curvature is finite.

Sharp generating contours will usually remain fixed on the object surface over time. (Exceptions occur only in the infrequent situations in which, due to changes in the line of sight with respect to the object, either sharp boundary becomes smooth or a flat face on one side of the generating point lines up with the line of sight.) Smooth generating contours will move along the surface of the object any time the surface orientation at a point fixed to the surface near the extremal boundary is changing with respect to the line of sight. Fig. 4 shows examples of both possibilities. The figure shows a view from above, with the sensor looking in the plane of the page and the objects rotating around an axis perpendicular to the line of sight. In Fig. 4(a), an object with a square cross section is being rotated. Fig. 4(b) shows an object with a circular cross section.

For sharp boundaries, a surface point close to a generating point in three-space projects onto the image at a location close to the image of the generating point. The surface point and the generating point move as a rigid body. For rigid body motion, differences in flow between the image of two points go to zero as the points become coincident in three-space. As a result, surface points arbitrarily close to the generating point project to the same flow values as the generating point itself.

For smooth boundaries, the situation is more complex. The surface points corresponding to the boundary may change over time, so that points on the surface near the generating point

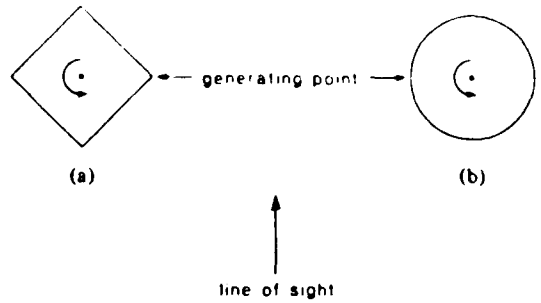


Fig. 4. (a) Generating contour at a sharp boundary remains fixed to the object surface. (b) Generating contour at a smooth boundary moves relative to the object surface.

and the generating point itself may not maintain a fixed relationship in three-space. The property described in (2) still holds for rotation in depth, however. The formal proof of this assertion is relatively complex and is given in Appendix B. (The Appendix actually shows that the limit of surface flow is equal to boundary flow for rotation of smooth objects around an arbitrarily oriented axis.) Informally, the result holds because the surface is tangent to the line of sight at the generating point, so that any motion of the generating point with respect to a point fixed to the surface is along the line of sight. The *difference* between the motion of the surface near the generating point and the motion of the generating point itself is a vector parallel to the line of sight and, hence, does not appear in the projected flow. This means that the motion of the boundary in the x direction will be the same as that of a point fixed to the surface at the instantaneous location of the generating point. The limit property holds because the surface flow varies continuously with x in the vicinity of the generating point, as long as we restrict our attention to points that are part of the same object.

To develop an algorithm for actually identifying the occluding surface at a detected boundary, we will start by assuming only translational motion is occurring. (Violations of this assumption are discussed below.) According to (2), we need only look at the flow at the edge point and immediately to either side to determine which side corresponds to the occluding surface. In practice, however, this in inadequate. Edges will be located imprecisely in each frame due to a variety of effects. This imprecision is compounded when the location of edge points is compared across frames to determine the flow of the edge. By considering the pattern of change in the Laplacian of the optical flow field, however, a simple binary decision test can be constructed to determine which surface velocity most closely matches that of the edge. As before, we will use a coordinate system with its origin at the location of some particular boundary point at a time t_0, the x axis oriented normal to the orientation of the boundary, and consider only $flow_x$, the projection of flow onto the x axis. In this new coordinate system, positive velocity values will correspond to motion to the right. We will assume that the flow field in the vicinity of the edge can be approximated by a step function. The algorithm developed here is unaffected by constants added to the flow field or by applying positive multiples to the magnitude of flow. Therefore, to simplify analysis, normalize the flow field by subtracting a constant value f_a such that the pro-

[illegible] of the two surfaces have equal magnitudes and opposite signs, and then multiply by a positive scale factor f_s such that the magnitudes will be normalized to 1 and -1 [i.e., $flow_x' = f_s(flow_x - f_a)$]. The resulting step edges can have one of two possible shapes, depending upon whether the surface to the left is, after scaling and normalizing, moving to the left or to the right (see Fig. 5).

When the two possible velocity functions are convolved with a Gaussian blurring kernel, the resulting functions are shown in Fig. 5(a) and (b). The Laplacian of these functions in the direction perpendicular to the edge is equal to the second derivative, and is shown in Fig. 5(c) and (d). These two cases may be described analytically as follows.

Case 1. Given the step function

$$s(x) = \begin{cases} 1, & x < 0 \\ -1, & x > 0 \end{cases} \tag{3}$$

convolve $s(x)$ with a Gaussian blurring function $g(x)$.

$$h(x) = g * s. \tag{4}$$

Let $s(x) = -2u(x) + 1$ where

$$u(x) = \begin{cases} 0, & x < 0 \\ 1, & x > 0. \end{cases}$$

Then

$$h(x) = 1 - 2\int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\lambda^2/2\sigma^2}\, d\lambda \tag{5}$$

$$h'(x) = \frac{2x}{\sigma^3\sqrt{2\pi}} e^{-x^2/2\sigma^2}. \tag{6}$$

Therefore,

$$h''(x) < 0 \quad \text{when} \quad x < 0$$

$$h''(x) > 0 \quad \text{when} \quad x > 0. \tag{7}$$

Case 2. The step function for case 2 is $-s(x)$, where $s(x)$ and $u(x)$ are defined above

$$h''(x) = \frac{-2x}{\sigma^3\sqrt{2\pi}} e^{-x^2/2\sigma^2}. \tag{9}$$

Therefore,

$$h''(x) > 0 \quad \text{when} \quad x < 0$$

$$h''(x) < 0 \quad \text{when} \quad x > 0. \tag{10}$$

At some later time t_1, the entire second derivative curve $h''(x)$ will have shifted right or left, depending upon whether the edge moves with the surface moving to the right or left. Based upon the analysis above, in case 1, if the left surface is occluding, the second derivative curve will be moving to the right and the sign at the origin will become negative, while if the right surface is occluding, the curve will be moving left and the sign at the origin will be positive. In case 2, if the left surface is occluding, the curve will be moving to the left and the sign at the origin will be negative; while if the right surface is occluding, the curve will be moving to the right and the sign at the origin will be positive. Note that in both cases, when the left surface is the occluding surface, the sign at the origin will become negative, and when the right surface is occluding, the sign at the origin will become positive. This is illustrated in Fig. 5(e) and (f). In the original, unrotated coordinate system, this is equivalent to stating that at time t_1 the direction normal to the edge for which the second directional derivative of optical flow is positive, evaluated at the location of the edge at t_0, points toward the occluding surface. (The approach is similar to that used in [9] to determine the direction of motion of an intensity contour.) This analysis may be extended to the general case where the original step function has not been normalized. The direction of the second derivative at t_1 must now, however, be evaluated at the point $(x_0, y_0) + (t_1 - t_0)f_a$. (As f_a is the average flow of the surfaces on either side of the boundary, this point may be thought of as lying half-way between the two possible image locations of the boundary at time t_1.)

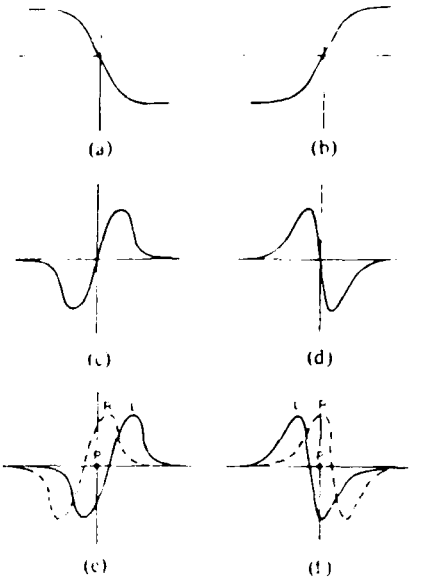


Fig. 5. Smoothed magnitude of flow for (a) case 1 and (b) case 2. (c) and (d) Laplacian of the functions in (a) and (b). (e) and (f) Two possible locations of the Laplacian curves after an interval of time. The dashed curve indicates the location of the curve if the edge moves with the surface to the right. The solid curve indicates the location of the curve if the edge moves with the surface to the left.

In practice, difficulties may arise for very large differential flows between the two surfaces. The second derivative function $h''(x)$ approaches zero away from the zero crossing. Noise sensitivity of the classification technique is likely to increase when the value is small. It is useful to determine a guideline for the size of the Gaussian blurring kernel to ensure that the curve will be observed near its extrema, where the sign is more likely to be correct. The form of the function $h''(x)$ may be simplified by substitution for analysis purposes. Let

$$b = \frac{x}{\sigma\sqrt{2}} \quad \text{and} \quad c = \frac{2}{\sigma^2\sqrt{\pi}}. \tag{12}$$

Then, in case 1,

$$h''(x) = f(b) = cbe^{-b^2} \tag{13}$$

$$f(b) = c e^{-b^2/2}(1 - 2b^2). \tag{14}$$

The extrema of $f(b)$ will occur at $b = \pm 1/\sqrt{2}$, and the extrema of $h''(x)$ occur at $x = \pm\sigma$. The ratio

$$\frac{h''(2.7\sigma)}{h''(\sigma)} = 0.12 \tag{15}$$

indicates that at $\pm 2.7\sigma$ the magnitude of $h''(x)$ is 12 percent of its magnitude at the extrema, and thus is relatively close to zero. If the noise is such that the sign will be accurate when the expected Laplacian value is at least 10 percent of the extrema value, then a Gaussian blurring kernel should be used of standard deviation at least 1/2.7 of the maximum expected magnitude of flow of the edge. For cases where the noise presents more of a problem, a Gaussian of larger standard deviation should be used. The analysis for case 2 can be performed similarly with the same result.

The algorithm is implemented as follows. Optical flow fields are obtained for two temporally adjacent image pairs. Approximation to the Laplacians of Gaussian blurred versions of these flow fields are calculated by computing the difference of the flow fields convolved with two different Gaussian kernels. (Again, the componentwise Laplacian is used.) As before, edge points are found in the first flow field by searching for vector reversals in the Laplacian of the field. At such points, the value of the smoothed flow field obtained from the larger of the Gaussian kernels is considered to approximate the average flow of the two surface regions on either side of the edge. This average flow is used to find the appropriate offset to add to the edge location to find P, a point midway between the two possible edge locations in the second Laplacian field. Next, the direction perpendicular to the edge point is estimated by finding the direction of greatest change in the Laplacian of the first flow field. The Laplacian of the second flow field at the point P is then examined. The Laplacian component in the second field perpendicular to the edge orientation points toward the occluding surface.

An example of this technique applied to an image sequence is shown in Fig. 6. The leopard translates from left to right approximately equally between frames 1, 2, and 3 in Fig. 6(a). The edge points shown in Fig. 6(b) are obtained as described in Section II. At each edge point, an offset based on the flow vector from the smoother version of the field at that point is added to the location of the edge point. The resulting location is examined in the Laplacian of the second flow field. The component of this Laplacian perpendicular to the edge will point toward the occluding surface. Shown in Fig. 6(c) are the edge points, each of which has an associated line segment. The line segment projects in the direction of the occluding surface, as determined by the algorithm. The correct classification is made for all except a few points at the bottom of the edge. In this region, several nearby tokens were matched in one frame pair but not the other, significantly affecting the smoothed flow fields in the neighborhood of the boundary.

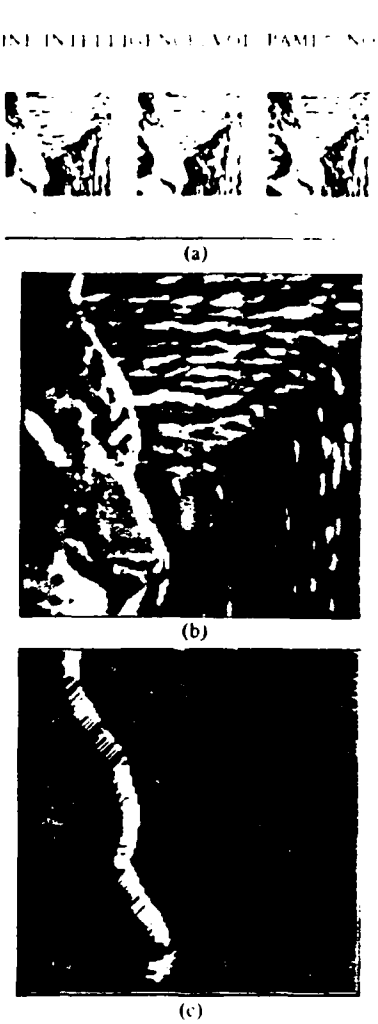


Fig. 6. (a) Image sequence. (b) Detected boundary overlaid onto first frame of sequence. (c) Identification of occluding surface. Each edge point has a line segment projecting from it toward the occluding surface.

IV. Rotational Motion

Rotation in depth introduces several complexities for the analysis of optical flow at occlusion boundaries. The first is an unexpected corollary of (2): in certain situations, there is *no* discontinuity in flow at occlusion boundaries. This occurs for pure rotation in depth of objects that are circularly symmetric, rotating about their axis of symmetry, and otherwise stationary with respect to the background. In such cases, the image of the boundary over time maintains a fixed position with respect to the background. As a consequence of (2), the projected surface flows on either side of the boundary are identical and are the same as the boundary flow itself. Fortunately, the zero-crossing-based boundary detection method is still usually applicable, although the detected location of the boundary may be displaced.

The second complication involves the determination of occluding surfaces. Rotations in depth produce a dynamic *self-occlusion*—the rotating object occludes sections of itself over time. In the situation described in the previous paragraph, self-occlusion is the only dynamic occlusion occurring. In these circumstances, the relationship in (2) is of no direct value in identifying the occluding surface. No information is available on which side of the boundary corresponds to a true occluding surface. (The situation is truly ambiguous in that two very different classes of spatial organizations can produce the same flow pattern.) If the rotating object is also translating relative to the background, if the object is not rotationally symmetric, or if it

is not rotating around an axis of symmetry, then (2) will, in principle, correctly identify the occluding surface. Difficulties arise in practice, however, because the algorithm given above depends on surface flow in the neighborhood of the boundary, not just at the edge. In the presence of rotation in depth, misclassifications are possible, particularly if no translation relative to the background is occurring and/or the rotating object is small, leading to rapidly changing flow values near the extremal boundary.

Rotation also complicates inferences about relative depth based on the analysis of occlusion boundaries. For translational motion, the occluding surface on one side of a boundary is necessarily in front of the occluded surface. For rotation in depth, the occluded and occluding surfaces are on the same side of the boundary, and no definitive information is available about the surface on the other side of the boundary. (Reference [10] shows an example in which a nonrotating surface on one side of a boundary is in front of a rotating surface on the other side of the boundary.) One approach to determining the actual relative depth involves first determining whether or not rotation in depth is actually occurring. Such as analysis is beyond the scope of this paper (see [11]). As an alternative, an analysis of surface regions that are appearing or disappearing due to dynamic occlusion gives information about the occluded surfaces at a boundary [10]. The method described here gives information about the occluding surface. By combining the two approaches, self-occlusion is recognized by noting a boundary where one side is marked as both occluding and occluded.

V. Conclusion

Motion-based boundary detection is sensitive only to depth discontinuities and/or object boundaries. Thus, unlike intensity-based edge detection, all detected edge points are of direct significance to the interpretation of object shape. On the other hand, significant edges will not be detected unless there is perceived motion between the surfaces on either side. Motion-based analysis offers another significant advantage. In most cases, the side of a boundary corresponding to the occluding surface can be identified. As we have shown, this is possible for general motion, not just for a sensor moving through an otherwise static environment. This determination is quite difficult using only static information, and has received only little attention (e.g., [12]).

Appendix A

The following is an analysis of the appropriateness of using zero crossings in the componentwise Laplacian of a flow field to detect contours of maximal rate of change in the flow field.

Theorem: Let V be a twice continuously differentiable vector field, let N be an open neighborhood containing the origin such that $\partial V/\partial y$ is constant on N, let L be the intersection of N and the y axis, and let u be a unit vector. Then $|\nabla V \cdot u|^2$ has an extremum in the x direction on L if and only if $u_x (u \cdot \nabla V) \cdot \nabla^2 V$ has a zero crossing on L.

Justification: The magnitude of the directional derivative in the u direction is

$$|\nabla V \cdot u|^2 = (\nabla V_x \cdot u)^2 + (\nabla V_y \cdot u)^2 \tag{16}$$

$$= \left[u_x \frac{\partial V_x}{\partial x} + u_y \frac{\partial V_x}{\partial y}\right]^2 + \left[u_x \frac{\partial V_y}{\partial x} + u_y \frac{\partial V_y}{\partial y}\right]^2 \tag{17}$$

$$= u_x^2 \left[\left[\frac{\partial V_x}{\partial x}\right]^2 + \left[\frac{\partial V_y}{\partial x}\right]^2\right] + 2u_x u_y \left[\frac{\partial V_x}{\partial x}\frac{\partial V_x}{\partial y} + \frac{\partial V_y}{\partial x}\frac{\partial V_y}{\partial y}\right] + u_y^2 \left[\left[\frac{\partial V_x}{\partial y}\right]^2 + \left[\frac{\partial V_y}{\partial y}\right]^2\right] \tag{18}$$

$$= u_x^2 \left|\frac{\partial V}{\partial x}\right|^2 + 2u_x u_y \frac{\partial V}{\partial x} \cdot \frac{\partial V}{\partial y} + u_y^2 \left|\frac{\partial V}{\partial y}\right|^2. \tag{19}$$

The partial derivative of this quantity can be simplified as follows:

$$\frac{\partial}{\partial x} |\nabla V \cdot u|^2 = 2u_x^2 \left[\frac{\partial V}{\partial x} \cdot \frac{\partial^2 V}{\partial x^2}\right] + 2u_x u_y \left[\frac{\partial V}{\partial y} \cdot \frac{\partial^2 V}{\partial x^2}\right] \tag{20}$$

$$= 2u_x \left[u_x \frac{\partial V}{\partial x} + u_y \frac{\partial V}{\partial y}\right] \cdot \frac{\partial^2 V}{\partial x^2} \tag{21}$$

$$= 2u_x (u \cdot \nabla V) \cdot \frac{\partial^2 V}{\partial x^2} \tag{22}$$

since $\partial V/\partial y$ is constant on N. For the same reason, $\partial^2 V/\partial y^2 = 0$ and $\partial^2 V/\partial x^2 = \nabla^2 V$. Therefore, $\partial/\partial x \, |\nabla V \cdot u|^2$ has a zero crossing whenever $u_x (u \cdot \nabla V) \cdot \nabla^2 V$ does. But $|\nabla V \cdot u|^2$ has an extremum in the x direction whenever $\partial/\partial x \, |\nabla V \cdot u|^2$ has a zero crossing. ∎

Whenever the Laplacian $\nabla^2 V$ has a zero crossing, so must $u_x (u \cdot \nabla V) \cdot \nabla^2 V$, except when $u_x (u \cdot \nabla V) = 0$, which is unlikely because real edges are places with steep gradients. Zero crossings in the Laplacian will therefore almost always correspond to extrema in the magnitude of the directional derivative, with respect to almost all directions. It is possible for the magnitude of the directional derivative to have an extremum without a zero in the Laplacian because the component at right angles to the preferred direction defined by $u \cdot \nabla V$ need not be small. If there is no variation of the field parallel to the edge, then the steepest directional derivative occurs in the direction normal to the edge; and if the variation parallel to the edge is much less than that normal to the edge, as we expect for most images, then the steepest directional derivative occurs in a direction nearly normal to the edge. If we choose u in the x direction, then $u \cdot \nabla V$ will be parallel to $\partial V/\partial x$, so that the above theorem states the component of the Laplacian in the direction parallel to the difference in the flow on both sides of the boundary will have a zero crossing. The Laplacian can fail to have a direction reversal at an edge only if the component of the Laplacian at right angles to the flow difference is not small, which occurs when the normal component of the flow gradient at an edge is changing in direction more rapidly

than it is changing in magnitude. Such situations do not appear to be common in real optical flows, and can occur only when the unfiltered flow is changing appreciably in a neighborhood of the edge for at least one of the two surfaces. For the case of a boundary between two surfaces with distinct uniform flows on each surface, the smoothed Laplacian has a directional zero crossing in all directions except along the boundary. In that direction, the value of the smoothed Laplacian is zero.

The extremum can be either a maximum or a minimum. A maximum is of course desired, and minima are discarded in practice by requiring the slope of the zero crossing to be sufficiently steep. While this is not a guaranteed test, it works in almost all cases because of the Gaussian filtering applied to the images before the Laplacian is calculated. Minima in the gradient usually correspond to areas where the field is uniform, and due to the tails on a Gaussian curve, gradients near the minima tend to be small, with small values for derivatives of all orders.

Appendix B

This Appendix contains the analysis showing that the limit of surface flow is equal to boundary flow for the rotation of smooth objects for orthographic projections. Any motion of a rigid body can be described by giving the trajectory of an arbitrary point attached to the body and the instantaneous rotation about some axis passing through that point. Define a set of Cartesian axes (X, Y, Z) with the origin at the distinguished point on the body and with the Z axis along the axis of rotation, and let (r, θ, ϕ) be spherical coordinates with respect to these axes. Let the orientations of the axes (X, Y, Z) be fixed with respect to the axes (x, y, z) of the image plane coordinates, so that the angular velocity of an arbitrary rotation is the same in both coordinate systems. Let the surface of the body be described by

$$r = R(\theta - \psi(t), \phi) \tag{23}$$

where $\psi(0) = 0$, so that $r = R(\theta, \phi)$ at time $t = 0$. The parameter $\alpha = \theta - \psi(t)$ is the longitudinal angle of a point fixed to the surface at $t = 0$, and points with constant values of α and ϕ rotate along with the surface. Since $\theta = \alpha + \psi(t)$, $\omega = d\psi/dt$ gives the angular velocity of the object about the Z axis.

At some particular instant of time, let G be a generating point (r_g, θ_g, ϕ_g) and $\boldsymbol{n}$ be the unit surface normal at G. Since G is a generating point and orthographic projection is involved, $\boldsymbol{n}$ will be parallel to the image plane. The normal component of the flow for an arbitrary point $\boldsymbol{p} = (r, \theta, \phi)$ fixed to the surface is as follows:

$$\begin{aligned} f_x(\boldsymbol{p}) &= (\boldsymbol{\Omega} \times \boldsymbol{p}) \cdot \boldsymbol{n} \\ &= \omega R(\theta - \psi, \phi) \sin\phi [-\boldsymbol{n}_X \sin\theta + \boldsymbol{n}_Y \cos\phi] \end{aligned} \tag{24}$$

where $\boldsymbol{\Omega}$ is the vector angular velocity of magnitude ω and oriented along the Z axis. The orientation of $\boldsymbol{\Omega}$ and $\boldsymbol{n}$ may be changing, but the analysis below is based on the instantaneous values of both quantities at some particular point in time.

The x axis in the image plane is oriented parallel to the constant unit vector $\boldsymbol{n}$. Since we are working with an orthographic projection, the x coordinate of the point $\boldsymbol{p}$ is as follows:

$$x = \boldsymbol{p} \cdot \boldsymbol{n} = R(\theta - \psi, \phi)\, [p(\theta, \phi) \cdot \boldsymbol{n}] \tag{25}$$

$$[p(\theta, \phi) \cdot \boldsymbol{n}] = \boldsymbol{n}_X \sin\phi \cos\theta + \boldsymbol{n}_Y \sin\phi \sin\theta + \boldsymbol{n}_Z \cos\phi \tag{26}$$

where p is the unit vector parallel to $\boldsymbol{p}$. Since the generating point is on the extremal boundary of the object, x must have an extremum at the generating point for variations in both θ and ϕ. This leads to

$$\frac{\partial x}{\partial \theta} = 0 = \frac{\partial R(\theta - \psi, \phi)}{\partial \theta} [p(\theta, \phi) \cdot \boldsymbol{n}] + R(\theta - \psi, \phi) \frac{\partial}{\partial \theta} [p(\theta, \phi) \cdot \boldsymbol{n}] \tag{27}$$

$$\frac{\partial x}{\partial \phi} = 0 = \frac{\partial R(\theta - \psi, \phi)}{\partial \phi} [p(\theta, \phi) \cdot \boldsymbol{n}] + R(\theta - \psi, \phi) \frac{\partial}{\partial \phi} [p(\theta, \phi) \cdot \boldsymbol{n}] \tag{28}$$

for $\theta = \theta_g, \phi = \phi_g$. Let x_g denote the x coordinate of the generating point. From (25), the flow of the boundary is as follows:

$$f_b = \frac{dx_g}{dt} = \left[\frac{d}{dt} R(\theta - \psi, \phi) [p(\theta, \phi) \cdot \boldsymbol{n}] \Big|_{\theta = \theta_g, \phi = \phi_g} \right] \tag{29}$$

$$\begin{aligned} &= \frac{\partial R(\theta - \psi, \phi)}{\partial \theta} [p(\theta, \phi) \cdot \boldsymbol{n}] \left[\frac{d\theta_g}{dt} - \frac{d\psi}{dt} \right] \\ &\quad + R \frac{\partial}{\partial \theta} [p(\theta, \phi) \cdot \boldsymbol{n}] \frac{d\theta_g}{dt} \\ &\quad + \frac{\partial R(\theta - \psi, \phi)}{\partial \phi} [p(\theta, \phi) \cdot \boldsymbol{n}] \frac{d\phi_g}{dt} \\ &\quad + R \frac{\partial}{\partial \phi} [p(\theta, \phi) \cdot \boldsymbol{n}] \frac{d\phi_g}{dt} \end{aligned} \tag{30}$$

evaluated at $\theta = \theta_g, \phi = \phi_g$. From (27), (28), and (26) we get

$$\begin{aligned} f_b &= -\frac{d\psi}{dt} \frac{\partial R(\theta - \psi, \phi)}{\partial \theta} [p(\theta, \phi) \cdot \boldsymbol{n}] \\ &= \frac{d\psi}{dt} R \frac{\partial}{\partial \theta} [p(\theta, \phi) \cdot \boldsymbol{n}] \end{aligned} \tag{31}$$

$$= \omega R(\theta_g - \psi, \phi_g) \sin\phi [-\boldsymbol{n}_X \sin\theta_g + \boldsymbol{n}_Y \cos\theta_g] \tag{32}$$

$$= f_x(0, 0) \tag{33}$$

using (24) and $d\psi/dt = \omega$. This establishes (2) for arbitrary orientations of the axis of rotation with respect to the image plane, assuming an orthographic projection.

Acknowledgment

M. Kaveh, D. Boley, and L. Qi provided much useful assistance in portions of the formal analysis.

References

[1] S. T. Barnard and W. B. Thompson, "Disparity analysis of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-2, pp. 333-340, July 1980.

[2] W. F. Clocksin, "Perception of surface slant and edge labels from optical flow: A computational approach," *Perception*, vol. 9, pp. 253-269, 1980.
[3] K. Nakayama and J. M. Loomis, "Optical velocity patterns, velocity sensitive neurons, and space perception: A hypothesis," *Perception*, vol. 3, pp. 63-80, 1974.
[4] H. C. Longuet-Higgins and K. Prazdny, "The interpretation of a moving retinal image," in *Proc. Roy. Soc. London*, vol. B-208, 1980, pp. 385-397.
[5] D. Marr and E. Hildreth, "Theory of edge detection," in *Proc. Roy. Soc. London*, vol. B-207, 1980, pp. 187-217.
[6] W. B. Thompson, K. M. Mutch, and V. A. Berzins, "Edge detection in optical flow fields," in *Proc. 2nd Nat. Conf. Artif. Intell.*, Aug. 1982.
[7] J. L. Crowley and R. M. Stern, "Fast computation of the difference of low-pass transform," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, pp. 212-222, Mar. 1984.
[8] P. J. Burt, "Fast filter transforms for image processing," *Comput. Graph. Image Processing*, vol. 16, pp. 20-51, 1981.
[9] D. Marr and S. Ullman, "Directional selectivity and its use in early visual processing," in *Proc. Roy. Soc. London*, vol. B-211, pp. 151-180, 1981.
[10] K. M. Mutch and W. B. Thompson, "Analysis of accretion and deletion at boundaries in dynamic scenes," *IEEE Trans. Pattern Anal Machine Intell.*, vol. PAMI-7, pp. 133-138, 1985.
[11] W. B. Thompson, K. M. Mutch, and V. A. Berzins, "Analyzing object motion based on optical flow," in *Proc. 7th Int. Conf. Pattern Recog.*, July 1984.
[12] A. P. Witkin, "Intensity-based edge classification," in *Proc. 2nd Nat. Conf. Artif. Intell.*, Aug. 1982.

William B. Thompson (S'72-M'75) received the Sc.B. degree in physics from Brown University, Providence, RI, in 1970, and the M.S. and Ph.D. degrees in computer science from the University of Southern California, Los Angeles, in 1972 and 1975, respectively.

He is currently an Associate Professor in the Department of Computer Science at the University of Minnesota, Minneapolis. Previously, he was a member of the Image Processing Institute at the University of Southern California. His primary research interest is in the area of computer vision with an emphasis on the development of techniques for perceiving spatial organization. In addition, he is a principal in the expert problem solving research group at the University of Minnesota.

Dr. Thompson is a member of the American Association for Artificial Intelligence and the Association for Computing Machinery.

Kathleen M. Mutch (S'80-M'83) received the M.S. and Ph.D. degree in computer science from the University of Minnesota, Minneapolis, in 1981 and 1983, respectively.

She is currently an Assistant Professor in the Department of Computer Science at Arizona State University, Tempe. Her research interests include time-varying image analysis and applications of computer vision.

Dr. Mutch is a member of the Association for Computing Machinery, SIGART, SIGCAPH, and the American Association for Artificial Intelligence.

Valdis A. Berzins (S'76-M'78) received the S.B. degree in physics and the S.M. and E.E. degrees in electrical engineering in 1975, and the Ph.D. degree in computer science in 1979, all from the Massachusetts Institute of Technology, Cambridge.

He is presently Assistant Professor of Computer Science at the University of Minnesota. His research interests include database support for computer aided design, software engineering, and image analysis.

Analysis of Accretion and Deletion at Boundaries in Dynamic Scenes

KATHLEEN M. MUTCH, MEMBER, IEEE, AND WILLIAM B. THOMPSON, MEMBER, IEEE

Abstract—In dynamic scenes, the presence of object boundaries is often signaled by the appearance or disappearance of occluded surfaces over time. Such regions of surface accretion or deletion can be found using matching techniques similar to those used to determine optical flow in an image sequence. Regions in one frame that are not adequately matched by any region in previous frames correspond to accretion. Regions that have no matches in subsequent frames correspond to deletion. In either case, an occlusion boundary is present. Furthermore, by associating accretion or deletion regions with a surface on one side of a boundary, it is possible to determine which side of the boundary is being occluded. This association can be based purely on visual motion—the accretion or deletion region moves with the same image velocity as the remaining visible surface to which it is attached.

Index Terms—Dynamic scene analysis, edge detection, occlusion, optical flow, segmentation.

I. Introduction

LOCATING object boundaries in images is an important but difficult problem. Intensity-based edge detection provides ambiguous or misleading boundary information in many situations, such as textured regions. Motion-based techniques can provide more reliable results in these cases. At object boundaries where occlusion occurs, surface regions will typically appear or disappear over time when motion is present. These regions of changing visibility may be used to indicate both object boundaries and the side of the boundary corresponding to the occluded surface.

At a typical object boundary, one surface will be blocking the view of another more distant surface. In the presence of motion, regions of the more distant surface will often either appear or disappear from view over time. Such regions are called areas of *accretion* or *deletion*, respectively. A similar situation arises in stereo vision, where a region of the more distant surface near an occlusion edge will be visible in one image of the pair but invisible in the other image. Thus, recognition of accretion/deletion regions is a means of locating object boundaries in image sequences. In addition, accretion and deletion regions will belong to the occluded surface, providing sufficient information to determine which of the two surfaces at a boundary is being occluded. To recover the information available from such regions, it is necessary to determine both how regions of accretion and deletion in the imagery may be identified, and what characteristics of such regions permit identification of the occluded surface.

This paper describes a scheme to locate regions of accretion and deletion, and to identify occluding surfaces at a boundary using these regions. A technique which matches image features in two frames is used to determine feature displacement on the image plane. Areas in the image with a high percentage of features which are unmatchable in a previous or subsequent image are identified as accretion or deletion regions, respectively. These regions indicate the presence of an occlusion boundary. Since the accretion/deletion region belongs to the occluded surface, it will be displaced on the image plane in the same fashion as nearby areas of that surface. The occluded surface is then identified by determining which of the two surfaces adjacent to the accretion/deletion region displays a similar displacement on the image plane. This identification combines information about accretion and deletion with optical flow to produce a description of the occlusion boundary more complete than any existing technique based purely on flow alone.

II. Previous Work

Several research efforts in computational vision have utilized motion information to recover object boundaries. The basic idea behind most motion-based approaches is that image plane motion, or optical flow, across the object surface will be constant or slowly varying, and discontinuities in flow will occur only at object edges. Previous approaches either search for discontinuities in the optical flow, or group together regions of similar flow. Nakayama and Loomis [1] propose a local, center-surround operator for detecting object boundaries in flow fields. Clocksin [2] shows that zero-crossings will occur at edge locations in the Laplacian of the magnitude of the optical flow field when an observer translates through an otherwise static environment. Thompson *et al.* [3], [4] demonstrate that the Laplacian is useful as an edge detector in the more general case of unconstrained motion. After obtaining point velocities by template matching, Potter [5] groups all points with the same velocity into single object regions. Similarly, Fennema and Thompson [6] use the spatial and temporal gradients of intensity to obtain point velocities, and then consider all points with similar velocities to be part of the same object. Thompson [7] develops a grouping scheme

Manuscript received April 30, 1984; revised October 22, 1984. Recommended for acceptance by Ruzena Bajcsy. This work was supported in part by the National Science Foundation under Grant MCS-81-05215, the Air Force Office of Scientific Research under Contract F49620-83-0140, and by Zonta International

K. M. Mutch is with the Department of Computer Science, Arizona State University, Tempe, AZ 85287.

W. B. Thompson is with the Department of Computer Science, University of Minnesota, Minneapolis, MN 55455.

0162-8828/85/0300-0133$01.00

based upon both intensity and velocity information. Regions of both identical intensity and identical velocity are formed, followed by merging of adjacent regions based upon similarities, or at least lack of conflict, in intensity and velocity. With the exception of Clocksin's work [2], these flow-based techniques are unable to provide any indication of the occluded surface at an edge.

Accretion and deletion are fundamental to motion analysis based on differencing [8], [9]. These techniques subtract one image from another and then use the presence of regions of significant difference to infer properties of object boundaries and motion. The approach is most effective when a reasonably homogeneous object is moving relative to a homogeneous background with different luminence. Covering and uncovering of the background leads to significant differences between frames, allowing boundaries to be located. Analysis of these difference regions over time can often be used to associate the difference region with adjacent, nonchanging areas of the image sequence and thereby identify which side of the boundary is being occluded. This scheme is intensity based, and suffers when intensity contrasts occur that are not related to object structure. A textured object which changes location on the image plane, for example, will produce many regions of intensity difference which are not accretion or deletion regions.

Only limited experimentation has been directed at the role of accretion and deletion in human perception. Kaplan [10] showed that patterns of accretion and deletion in fields of moving random dots provide sufficient information for the judgment of relative depth by human subjects. In his stimuli, a single edge separated two regions of random dots, where each region moved coherently. The edge was implicit, being the line along which accretion and/or deletion occurred, and thus was not visible if all of the dots were stationary. Subjects consistently perceived the more distant surface to be the one which was undergoing accretion or deletion at a greater rate. This was true even when the implicit edge moved with a velocity different than the velocity of points on either surface. In these cases of inconsistent edge motion, there was more ambiguity in the perceptions of subjects, although the statistically significant perception was that the surface with a greater rate of accretion or deletion was more distant. This suggests that both edge velocity and accretion/deletion are important factors in the perception of depth at an edge, but that accretion/deletion information may be dominant.

III. Detecting Accretion/Deletion Regions

A motion-based scheme for identifying accretion and deletion regions is developed here. To recover motion on the image plane, corresponding structures in each frame of an image pair are located. The result of this is a *disparity vector field*, where each vector represents the change in image plane location of a structure. (Disparity is the discrete representation of optical flow arising from image sequences that are discretely sampled in time.) This correspondence is accomplished by *token matching*. A token is a distinctive region in the image which is identified by some predefined local operator. A set of tokens is obtained for each image in the pair, and an organized search is performed to match tokens from the first image to corresponding tokens in the second image using the relaxation labeling technique described in [11]. Possible matches between tokens in the two frames are evaluated based on two criteria: the similarity between properties of the tokens, and a surface smoothness measure that favors matches with disparities similar to neighboring tokens. An important aspect of this particular matching technique is that it can determine that a token in one frame is *unmatchable* if no token in the other frame satisfies the appropriate matching criteria. By basing the analysis on the motion of tokens in the image, many of the intensity contrast problems of a differencing system are circumvented.

Fig. 1. Location of an accretion/deletion region relative to the boundary indicates the direction of the occluded surface. In both cases shown above, the vertical line represents a boundary and the shaded area represents an accretion or deletion region. The arrow points toward the occluded surface.

Regions of accretion and deletion are identified by analyzing unmatchable tokens in either image. A token may not be matchable either because the token detector failed to find the corresponding structure in the other image of the pair, or because the corresponding token is not visible in the other image. Regions with a high ratio of unmatchable tokens to total tokens are likely to be regions of accretion or deletion. This motion-based, token-matching approach is an implementation of Kaplan's model for detecting such regions [12]. Kaplan argues that accretion and deletion are detected in the human visual system by isolating clusters of elements of optical texture, tracking them over time, and responding when they change in some way that is not topologically permissible. Token identification is equivalent to isolating elements of optical texture; token matching serves the purpose of tracking such elements over time; and analyzing unmatchable tokens is a response to some change which may be due to appearance or disappearance of a region.

IV. Identifying Occluded Surfaces

Not only can accretion/deletion patterns be used to locate boundaries, they provide information that allows the identification of the side of the boundary being occluded. Such information is beneficial when interpreting dynamic scenes. Several specific approaches are possible, though all are based on associating the accretion or deletion region with a surface on one side of the boundary. That surface is the one being occluded. One approach relies upon the relative location of the accretion/deletion region with respect to the precise position of the image of the boundary. This boundary is the actual point of occlusion, the accretion/deletion region being on the same side of the boundary as the occluded surface. Fig. 1 illustrates this concept. The primary difficulty in this approach is identifying the boundary location relative to the accretion/deletion region. In particular, motion-based edge de-

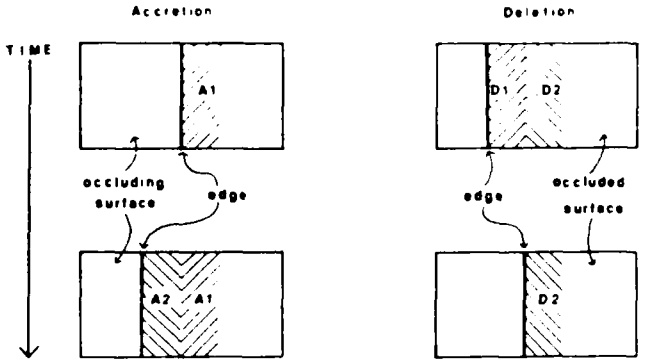


Fig. 2. The location of new accretion/deletion regions relative to previous such regions indicates the direction of the occluded surface. The second accretion region appears to the opposite side of the first accretion region as the occluded surface. The second deletion region occurs to the same side of the first deletion region as the occluded surface.

tection cannot locate the boundary precisely enough without first knowing which surface is occluded. The inadequacies of intensity-based edge detectors for this purpose are well known, particularly when applied to textured surfaces.

An alternative approach involves identifying the location of an accretion or deletion region relative to the location of such a region at a previous instant in time [13]. New accretion regions will appear to the side of previous accretion regions opposite the remainder of the occluded surface. New deletion regions will occur on the same side of previous deletion regions as the occluded surface (see Fig. 2). A disadvantage of this approach is the necessity to track and locate whole accretion/deletion regions over time.

The approach for identifying occluded surfaces from accretion/deletion regions which is developed in this paper requires the recognition of similarities between such regions and one of the two surfaces on either side of the boundary. Since the accretion/deletion region belongs to the occluded surface, it will share certain properties with that surface. The common property could be intensity or texture, although the problems inherent in most intensity-based analyses make these alternatives undesirable. Once again, motion-based properties may be more reliable. One such property is the disparity of tokens on the image plane. Disparity varies slowly over the surface of almost all rigid objects. Accretion or deletion tokens will thus exhibit disparities which are nearly identical to nearby token disparities on the same surface, while token disparities on different surfaces will usually vary.

V. Implementation

The system which was developed to detect occluded surfaces from regions of accretion or deletion uses token matching to obtain disparity vector fields. Unmatched tokens in clusters of high density are classified as accretion or deletion tokens, depending upon whether they have matches in subsequent or previous frame pairs. The disparity of accretion tokens after their appearance, or of deletion tokens prior to their disappearance, is obtained. Nearby tokens which are not accretion or deletion tokens and which have known disparities are identified and are used to identify the surface to which the accretion or deletion tokens belong. Such tokens with similar disparities to an accretion or deletion point lie on the occluded surface.

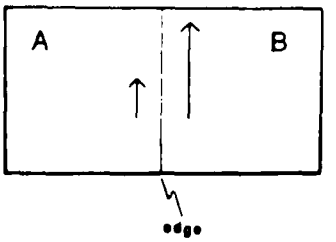


Fig. 3. The optical flow of surfaces *A* and *B* is indicated by the vectors on those surfaces. Since neither surface exhibits any flow perpendicular to the edge, there will be no accretion or deletion regions.

Three frames in an image sequence are required. Disparity fields D1 and D2 are obtained for frames 1 and 2, and for frames 2 and 3, respectively. Accretion points are not visible in frame 1, but do appear in frames 2 and 3. Tokens first appearing in frame 2, and thus having no associated match in frame 1, are noted. If these tokens have a match in frame 3, and if they are in a region with a high ratio of such tokens to total tokens, they are considered to be points of accretion. The disparity of accretion points is provided by D2. For every accretion point, a search is made within a neighborhood about the point location in frame 2. Tokens which are matched in D2, but which are not marked as accretion points are found. All of these tokens which have disparities similar to the accretion point are considered as a cluster. A vector pointing towards the center of the cluster is assigned to each accretion point, and indicates the direction from that point to the occluded surface.

Deletion points are visible in frames 1 and 2, but not frame 3. Tokens which are indicated as unmatchable in frame 2 are noted. If these tokens have a match in frame 1 and if they are in a region with a high ratio of such tokens to total tokens, they are considered to be points of deletion. The disparity of deletion points is provided by D1. For every deletion point, a search is made within a neighborhood about that point location in frame 1. Tokens which are matched in D1, but which are not marked as deletion points, are found. All of these tokens which have disparities similar to the deletion point are considered as a cluster. As before, a vector in the direction of the center of the cluster is assigned to each deletion point and indicates the direction from that point to the occluded surface.

VI. Limitations

This boundary detection technique requires a moderately dense token set, both to find accretion/deletion regions, and to determine image-plane displacements. This means that the two surfaces adjacent to the edge must be distinctly textured. In addition, there must be some component of optical flow perpendicular to the occlusion boundary, or neither accretion nor deletion will occur. In particular, motion exactly parallel to the boundary will produce no accretion or deletion regions (see Fig. 3). Perspective viewing of translating objects in principle leads to difficulties similar to those associated with rota-

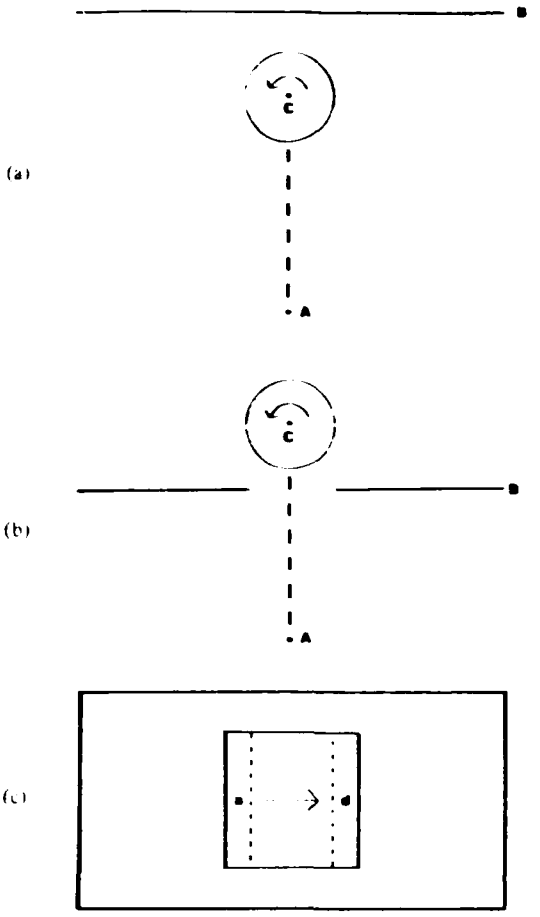


Fig. 4. (a) Overhead view of a cylinder rotating counter-clockwise about an axis at C, in front of a stationary background B. The viewer is at A, and the line of sight is along the dotted line. (b) The rotating cylinder seen through an aperture in surface B which is now in front. (c) When either (a) or (b) are viewed from point A, accretion regions a will occur along the left edge on the cylinder, deletion regions d along the right edge. While the cylinder is correctly identified as the occluded surface, there is insufficient information to determine the relative depth between the cylinder and the surface at B.

tion in depth (see below), as the perspective effects can be locally described as a combination of rotation and scale change. Fortunately, the practical difficulties caused by perspective effects are minimal. When objects are translating in front of a background, the size of accretion/deletion regions due to translation is almost always much greater than accretion/deletion regions that appear due to effective rotation of the object.

Certain rotations lead to potentially confusing situations when analyzing occlusion boundaries. Fig. 4(a) shows an overhead view of a cylinder rotating in depth. Fig. 4(c) shows the accretion/deletion regions that arise if there is no relative motion between the cylinder and the background surface. The analysis above assigns the accretion and deletion regions to the cylinder. Thus, the cylinder, not the background surface, is indicated as the surface being dynamically occluded. This is the correct interpretation, as the rotation in depth causes the cylinder to occlude itself over time. However, while the dynamic occlusion is correctly recognized, no information

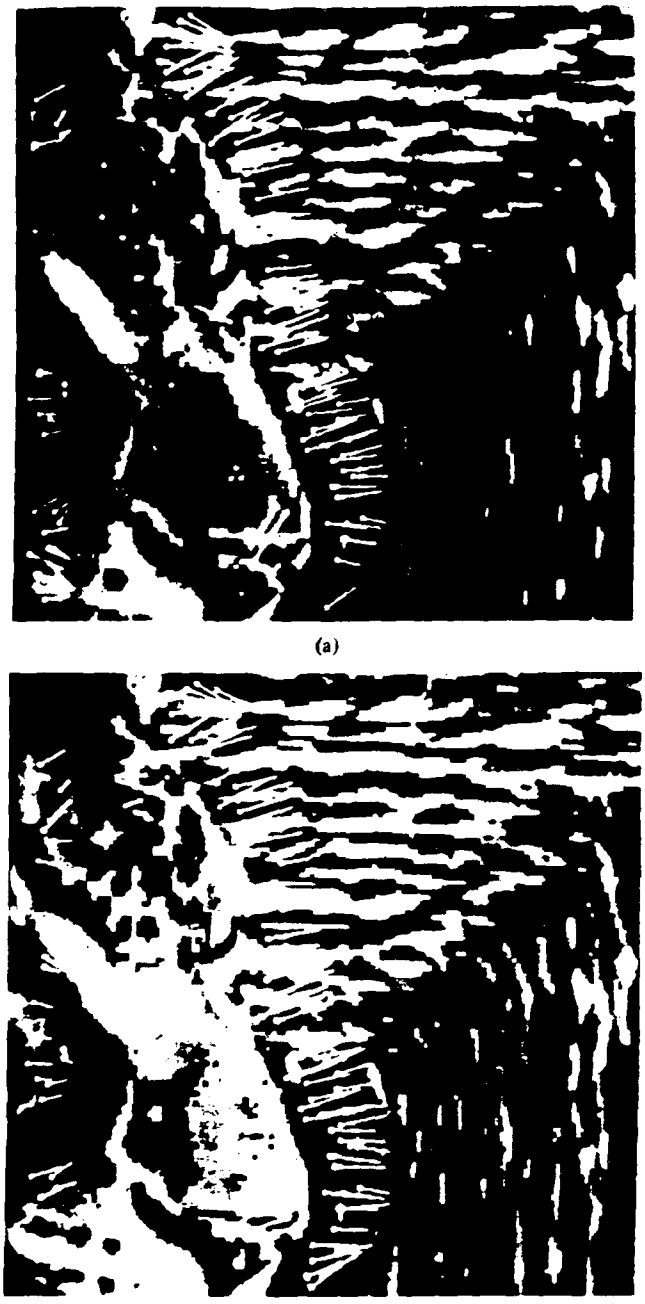
Fig. 5. Image sequence in which the leopard is translating from left to right.

is directly available about the relative depths to the surfaces on either side of the boundary. In fact, it is possible that the surrounding surface in the image is actually in front of the cylinder [Fig. 4(b)], yet generates the same image sequence.

A different complication occurs if the rotating object is moving with respect to the background surface, the cross section of the object is not circular, or the object is not rotating about its axis of symmetry. In all of these situations, accretion and/or deletion will be occurring on both sides of the actual boundary. The method given above is still valid and will identify both sides of the boundary as occluded surfaces. The problem again arises when trying to infer relative depth given an identification of the occluded surface. The determination of relative depth at a dynamic occlusion boundary when rotation is occurring is made possible by combining accretion/deletion analysis as described in this paper with an optical flow based approach [4]. This second technique uses the relationship between the flow of a boundary and the surface flows on either side of the boundary to identify occlud*ing* surfaces. Accretion/deletion analysis locates occlud*ed* surfaces. When taken together, both the occlusion of one surface by another and the self-occlusion resulting from rotation in depth can be recognized and appropriately interpreted.

VII. Example

The system implementation described above was applied twice to the image sequence shown in Fig. 5, first processing the sequence in the order shown, then in the reverse order. All images had a resolution of 128 × 128. There were approximately 1000 tokens identified in each image, and over 800 of these were matched in every image pair. As is usual with token matching systems, the density of tokens (and thus disparity vectors) varied across the image, being higher in areas of fine texture. An 11 × 11 square neighborhood, centered at the unmatched point, was used for computing the density of unmatched tokens. This size was chosen to be small enough so that most of the neighborhood fell within the accretion/deletion region, yet big enough to contain a reasonable number of tokens (usually 6 to 12). If 80 percent of the tokens in this neighborhood were unmatched in the same way as the point under consideration, then the point was labeled "accretion" or "deletion." This ratio was chosen to be selectively high, and yet to allow for some incorrect matches in the neighborhood, or some extension of the neighborhood out of the accretion/deletion region. A 31 × 31 window, centered at the accretion/deletion point, was searched to find clusters of similar disparity vectors. This size was chosen to be large enough to include portions of both surfaces outside the accretion/deletion region, yet not so large as to extend beyond these sur-



(a)

(b)

Fig. 6. (a) Results of occluded surface determination based upon accretion/deletion regions for the sequence of three frames in Fig. 5. Square white boxes are locations of accretion or deletion points in frame 2. The line emanating from each box points in the direction of the occluded surface. (b) Results of occluded surface determination when the sequence of three frames in Fig. 5 is processed in the reverse order.

faces. Disparity vectors were considered "similar" if they differed by no more than 2 pixels in each of the x and y components. The actual values of most of these parameters will, in general, depend upon factors such as the resolution of the images, the amount of texture, and the maximum expected disparity.

The results of processing in the forward direction are shown in Fig. 6(a). All of the square white points represent accretion or deletion frame 2 tokens, which were matched in the second (D2) or first (D1) disparity field. The line which emanates from each box projects toward the surface which the algorithm indicates is being occluded. The set of tokens to the right of the leopard are deletion points. Tokens near the left border of the image are accretion points, which appear as more of the leopard moves into the field of view. Vectors associated with these points indicate that the leopard is being occluded by the surrounding frame. Except for six noise points, all accretion and deletion tokens have an associated vector pointing in the correct direction. The noise points are not in the accretion or deletion regions, but rather occur in or near untextured regions, or on the edge of the accretion/deletion regions. As a result, there are either no other tokens in the vicinity, or else a large number of unmatched tokens in the neighboring accretion/deletion region. These points are thus incorrectly identified as accretion or deletion points.

Fig. 6(b) shows the results when the image sequence of Fig. 5 is processed in the reverse order. The disparity field D1 is now the set of matches for frames 3 and 2, and D2 for frames 2 and 1. Tokens to the right of the leopard are now accretion points, and tokens near the left border of the image are deletion points. Once again, except for nine noise points, all vectors correctly point toward the occluded surface. The noise points are due to the same causes described in the previous paragraph.

REFERENCES

[1] K. Nakayama and J. M. Loomis, "Optical velocity patterns, velocity sensitive neurons, and space perception: A hypothesis," *Perception*, vol. 3, pp. 63-80, 1974.

[2] W. F. Clocksin, "Perception of surface slant and edge labels from optical flow: A computational approach," *Perception*, vol. 9, pp. 253-269, 1980.

[3] W. B. Thompson, K. M. Mutch, and V. A. Berzins, "Edge detection in optical flow fields," presented at 2nd Nat. Conf. Artificial Intell., Aug. 1982.

[4] —, "Dynamic occlusion analysis in optical flow fields," Comput. Sci. Dep., Univ. Minnesota, Tech. Rep. 84-6, 1984.

[5] J. L. Potter, "Scene segmentation using motion information," *Comput. Graphics Image Processing*, vol. 6, pp. 558-581, Dec. 1977.

[6] C. L. Fennema and W. B. Thompson, "Velocity determination in scenes containing several moving objects," *Comput. Graphics Image Processing*, vol. 9, pp. 301-315, Apr. 1979.

[7] W.B. Thompson, "Combining motion and contrast for segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-2, pp. 543-549, Nov. 1980.

[8] R. Jain, W. N. Martin, and J. K. Aggarwal, "Segmentation through the detection of changes due to motion," *Comput. Graphics Image Processing*, vol. 11, pp. 13-34, Sept. 1979.

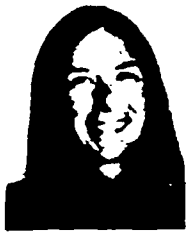
[9] R. Jain, "Extraction of motion information from peripheral processes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-3, no. 5, pp. 489-503, 1981.

[10] G. A. Kaplan, "Kinetic disruption of optical texture: The perception of depth at an edge," *Perception Psychophys.*, vol. 6, pp. 193-198, 1969.

[11] S. T. Barnard and W. B. Thompson, "Disparity analysis of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-2, pp. 333-340, July 1980.

[12] G. A. Kaplan, "Kinetic disruption of optical texture: The perception of depth at an edge," Ph.D. dissertation, Cornell Univ., Ithaca, NY, 1968.

[13] W. B. Thompson and S. T. Barnard, "Low-level estimation and interpretation of visual motion," *Comput.*, Aug. 1981.



Kathleen M. Mutch (S'80-M'83) received the M.S. and Ph.D. degrees in computer science from the University of Minnesota, Minneapolis, in 1981 and 1983, respectively.

She is currently an Assistant Professor in the Computer Science Department at Arizona State University, Tempe, AZ. Her research interests include time-varying image analysis and applications of computer vision.

Dr. Mutch is a member of the Association for Computing Machinery, SIGART, SIGCAPH, and the American Association for Artificial Intelligence.

William B. Thompson (S'72-M'75) received the Sc.B. degree in physics from Brown University, Providence, RI, in 1970, and the M.S. and Ph.D. degrees in computer science from the University of Southern California, Los Angeles, in 1972 and 1975, respectively.

He is currently an Associate Professor in the Computer Science Department of the University of Minnesota, Minneapolis, MN. Previously, he was a member of the Image Processing Institute at the University of Southern California. His primary research interest is in the area of computer vision, with an emphasis on the development of techniques for perceiving spatial organization. In addition, he is a Principal in the expert problem solving research group at the University of Minnesota.

Dr. Thompson is a member of the American Association for Artificial Intelligence and the Association for Computing Machinery.

END

1-87

DTIC